BLF

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 December 2000 (28.12.2000)

PCT

(10) International Publication Number
WO 00/79267 A2

(51) International Patent Classification[7]: G01N 33/48

(21) International Application Number: PCT/GB00/02446

(22) International Filing Date: 22 June 2000 (22.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

9914589.8	22 June 1999 (22.06.1999)	GB
0008161.2	3 April 2000 (03.04.2000)	GB

(71) Applicant *(for all designated States except US)*: SCHOOL OF PHARMACY, UNIVERSITY OF LONDON [GB/GB]; 29-39 Brunswick Square, London WC1N 1AX (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: NIZETIC, Dean [HR/GB]; School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX (GB). GROET, Jürgen [DE/GB]; School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX (GB).

(74) Agent: GILL JENNINGS & EVERY; Broadgate House, 7 Eldon Street, London EC2M 7LH (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/79267 A2

(54) Title: TREATMENT OF CANCER

(57) Abstract: The product of a gene located at human chromosome 21q11-21 has ubiquitin specific protease activity and binds to ubiquitin, polyubiquitin, ubiquitin-like protein SUMO-3, and proteins which are implicated in the repair/excision of DNA. It is believed to have a role in regulating cell growth, cell growth arrest and/or apoptosis and is implicated in carcinoma growth. The gene has been sequenced and cloned into a microorganism and the product's effect on cell growth is to be investigated. The gene has GenBank accession number AF134213 and has the HUGO-approved name USP25.

Treatment of Cancer

The present invention relates to the use of the product of the USP25 gene, which appears to have ubiquitin specific protease activity and may have protease activity on ubiquitin-like proteins, located at human chromosome 21q11 in the treatment, prophylaxis or diagnosis of cancer, especially solid tumours, more particularly lung cancer.

Ubiquitin mediated proteolysis by the 26S proteasome is responsible for the physiological regulation of the levels of many proteins which are key cell cycle and cell growth regulators (Hochstrasser 1995), as well as known tumour suppressors (Lane 1998).

Ubiquitin is a 76 amino acid polypeptide present in all eukaryotic cells, and highly conserved in evolution. Ubiquitin is conjugated to a target protein through an isopeptide bond between the ε-amino group of Lys in a target protein (ubiquitination), a process mediated by three groups of enzymes: ubiquitin activating enzymes (E1), ubiquitin conjugating enzymes (E2) and ubiquitin ligases (E3). Ubiquitinated proteins exist in a monoubiquitinated form, or a multiubiquitin chain: the former is not a degradation signal, while the latter, Lys-48-linked ubiquitin-ubiquitin(n) conjugate, works as a strong degradation signal when joined to a Lys in a target protein. Protein conjugated to polyubiquitin is then very rapidly and efficiently degraded by non-lysosomal, ATP-dependent degradation by the 26-S proteasome. The de-ubiquitinating enzymes (DUB's) broadly fall into two classes; ubiquitin specific proteases (USPs) and ubiquitin-C-terminal hydrolases (UCHs) are each capable of de-conjugating the ubiquitin-ubiquitin and ubiquitin-protein links, thereby converting polyubiquitin into mono-ubiquitin, and de-coupling (deubiquitinating) ubiquitin from the target protein, with the result of preventing the degradation of the target protein.

The class of UCH enzymes tends to include relatively small proteins (about 35-40kD) which have low specifity for ubiquitinated proteins. BAP1 is a UCH but is unusual in that it is larger, having a weight of nearly 100kD.

Several such enzymes are known, the sequences of which show some sequence homologies especially in two domains, the Cys and His domains.

Very broadly, two main functions have been observed among the various members of the USP (UBP) superfamily (Wilkinson 1997, Wilkinson and Hochstrasser 1998). The first is the generation of free ubiquitin from precursor fusion proteins or from peptide-linked polyubiquitin after proteolysis of the targeted protein, and the second is de-ubiquitination. When a protein is targetted for ubiquitin mediated degradation, it is linked to ubiquitin via an isopeptide bond between the C-terminus of ubiquitin and a lysine ε-amino group(s) of the acceptor protein. Once the conjugate is formed, it can have only two fates: non-lysosomal proteolysis mediated by the 26S proteasome resulting in total protein degradation, or de-conjugation from ubiquitin (de-ubiquitination), resulting in the rescue of the target protein from degradation (Wilkinson 1997, Wilkinson and Hochstrasser 1998).

The regulation of p53 levels by USP25, and consequential prescribed increased cellular predisposition to apoptosis (programmed cell death) upon overexpression of USP25, could be a mechanism by which neuronal loss occurs in Alzheimer's Disease associated with Down Syndrome. Some mechanism could be invoked in explaining the sporadic AD if overexpression of USP25 occurs through a somatic change in aging neurons.

The potential USP catalysed de-ubiquitination and rescue from degradation of tumour suppressors such as p53, could perhaps explain the link to deletions of USPs in solid tumours. De-ubiquitination could play a major role in the Mdm2 mediated control of p53 levels and its activation mechanism, since the ubiquitin-mediated proteasome degradation of p53 is an important effector arm of this recently revealed control pathway (Haupt et al., 1997, Kubbutat et al., 1997, Lane 1998). Halving the dose of a USP could give a pre-cancerous cell a selective advantage in proliferation, by diminishing its rate of "re-cycling" of p53, making it more difficult to achieve a threshold concentration of p53, necessary for its activation. The same hypothetical, gene-dose regulated, tumour suppression model could also

explain why trisomy 21 (Down's syndrome DS) seems to confer protection from development of solid tumours. Among DS children, neuroblastomas and nephroblastomas, and among DS adults, gynaecologic, digestive and breast cancers have very rarely been reported, and are significantly under-represented compared to the age-matched euploid population (Oster et al., 1975, Satge et al., 1998).

Baker *et al*, 1999 also propose a role for USP's in tumour formation and cell growth.

In recent years a number of other protein modifying polypeptide tags have been identified. Many of these are related to ubiquitin and have high levels of identity and similarity (determined using the BLAST algorithm, for instance) to ubiquitin itself. There is a recognised super family of such proteins which have been termed ubiquitin-like proteins (UbL) (Gong et al. 1997, Schwarz et al. 1998). The yeast Smt3 and human SUMO-1 (PIC1, Sentrin, hSmt3C), SUMO-2 (hSmt3A) and SUMO-3 (hSMT3B) belong to the same family of UbL proteins with approximately 50% identity between themselves, and some 15-30% identity and 40-60% similarity in amino acid sequence to ubiquitin (Lapenta et al. 1997, Mannen et al. 1996, Kamitani et al. 1998, Saitoh and Hinchey, 2000). Yeast and human UBC9 are capable of conjugating equally yeast or human UbL-s, but not ubiquitin (Schwarz et al. 1998). The SUMO-1,-2 and -3 have the C-terminal glycine, necessary for ubiquitination of the target protein's lysine residue, but unlike ubiquitin, do not have the Lys48 residue necessary for the formation of polyubiquitin chains through isopeptide bonds, which are the signal for the proteasome degradation (Saitoh and Hinchey, 2000). Nevertheless, yeast Smt3 protein can rescue the mutant Mif2 phenotype, a deficient centromere binding protein resulting in chromosome missegregation (Meluh and Koshland 1995). SUMO-1, as well as SUMO-3 (and probably also SUMO-2) are all capable of being attached by UBC9 to RanGap1, a Ran GTP-ase activating protein (Kamitani et al. 1998). This ATP-dependent attachment is essential for the binding between modified RanGap1 and RanBP2 binding protein, in

order to form functional nuclear pore complex, which controls export and import of molecules through the nuclear envelope (Mahajan et al. 1997, Matunis et al. 1998, Lee et al. 1998). In addition, UbL small proteins have been shown to modify the death domains of Fas (Okura et al. 1996), Tumour necrosis factor receptor1 (Okura et al. 1996), PML (a tumour suppressor implicated in the pathogenesis of acute promyelocytic leukaemia) (Kamitani et al. 1998b) and Rad51/52 DNA repair proteins (Shen et al. 1996a). Their conjugating enzyme, UBC9, has been shown to interact by Y2H technique with RAD51/52 DNA repair proteins, and the master tumour suppressor p53 (Shen et al. 1996b). Another UbL is NEDD8 (Kamitami *et al* 1997).

UbL's are conjugated and cleared from their targets by enzymes. Several UbL hydrolase enzymes have been identified which convert precursor UbL to active UbL. Some such enzymes interact with ubiquitin itself as well as with other UbL's. Proteases involved in cleavage of conjugates of UbL with target protein have been identified for instance SENP1 and SUSP-1, which were recently cloned (Kim et al. 2000, Gong et al. 2000a), and found to specifically cleave SUMO-1,-2 and -3, but not ubiquitin and NEDD8. The first human enzyme with classical USP structure (Cys, His domains) for which dual specificity to both ubiquitin and ubiquitin like protein was demonstrated was very recently published USP21 on chromosome 1q21 (Gong et al. 2000b). However, opposite from SENP1 and SUSP-1, this enzyme cleaves ubiquitin and Nedd8, but not SUMO-1,-2 or -3 (Gong et al. 2000b).

The proximal third of the chromosome 21 long arm is an exceptionally gene-poor region of the human genome as estimated by a number of criteria (Shimizu et al., 1995, Yaspo et al., 1995, Gardiner 1996) the estimates of gene-density range from one gene in a megabase to one gene in six megabases of genomic DNA. Until recently only three full length genes had been mapped in this region: STCH (a member of the hsp70 family) (Brodsky et al., 1995), RIP140 (protein functionally interacting with a variety of nuclear receptors such as estrogen receptor), (Cavailles et al., 1995) and ANA (a

member of the Tob/BTG1 family of tumour suppressors), (Kohno et al., 1998). This region is also an example of extremely highly methylated regions in the human genome.

Groet *et al.* (1998) describe a high-resolution bacterial contig map of 3.4 Mb of genomic DNA in human chromosome 21q11-q21, encompassing the region of elevated disomic homozygosity in Down's syndrome - associated abnormal myelopoiesis and leukemia, and which has shown a strong association with Alzheimer's disease (AD). It was suggested that the high resolution bacterial clone overlap map should be the basis for deriving a more complete transcriptional map of that region of the chromosome. It was hoped that this would lead to an explanation of the chromosome 21q11 linkage in familial early onset AD (FEOAD) families. In particular it was suggested that a modifier gene in that region could act together with the presenilin-1 gene to generate or modify the AD phenotype.

Further work by the present inventors has revealed a new gene in the proximal third of chromosome 21 and that the product of this gene has ubiquitin specific protease properties. It is postulated that the gene product and USPs generally may have a role in AD. The work has been published in Groet *et al* (2000) after the first priority date of the present application.

Valero, *et al.* (1999) published after the first priority date of the present application, have, in parallel identified this gene and pointed out the gene product's sequence homologies to known USP's in the conserred peptide domains previously identified e.g. by d'Andrea *et al* (1998). They postulate a role in Alzheimer's disease. This protein has the HUGO approved name USP25.

According to the present invention there is provided the new use of the product of the USP25 gene located at human chromosome 21q 11-21, a non-human homologue thereof or a functional fragment thereof in the manufacture of a medicament for use in the diagnosis, prophylaxis or treatment of cancer.

The canc r is usually a solid cancer, most often non-small cell lung carcinoma, or skin cancer.

A functional fragment herein means a protein having ubiquitin specific protease activity or UbL specific protease activity and comprising a portion with sequence homology with the product of the said USP25 gene.

It is believed that the ubiquitin specific protease activity of the protein having sequence ID 1 is responsible for its implication in the pathogenesis of cancer. USP activity may be determined using the technique described in the Examples below, in which using bacteria cotransformed with the USP gene and with a reporter gene encoding a fusion protein which is a ubiquitin-conjugated detectable protein. The protein may be an enzyme detectable by direct enzyme reaction, by enzyme-linked immune assay techniques, by autoradio-graphically or by direct staining after gel separation under conditions suitable to separate ubiquitin and cleaved protein from fusion protein.

From experiments conducted to determine with which proteins the product having sequence ID 1 interacts, we have found that there is interaction with ubiquitin, polyubiquitin and various ubiquitin precursors, as well as HHR23A (Matsutani *et al* 1994, GenBank Accession No D21235). There is also interaction with other ubiquitin-like proteins and with proteins which are known to interact with ubiquitin-like proteins, such as Sumo-3 (Mannen *et al* 1996, Kamitani *et al* 1996, Saitoh *et al* 2000, GenBank Accession No. NM 006937) and ubiquitin-like-specific conjugating enzyme 9 (Schwarz *et al* 1998, Lee *et al* 1998, GenBank Accession No. U 66867). The isolated protein may therefore be characterised further by having a positive interaction in a yeast-two-hybrid procedure with one or more, preferably all three, proteins having the sequences of GenBank Accession Nos. D 21235, NM 006937 and U 66867. Ubiquitin-like specific protease activity may be determined using techniques analogous to those used to determine ubiquitin specific protease activity, by using a substrate which is a fusion protein of the ubiquitin-like protein of interest and a detectable protein, and using the

usual separation and immune based or autoradiographic identification techniques.

The gene may be transcribed and translated within a cell line selected as positive for the native gene or an active (in terms of ubiquitin or UbL specific protease activity) fragment thereof. The gene has preferably been introduced into a microorganism or a cell line in a form in which it can be transcribed and translated and the microorganism or the cell-line, as the case may be, has been cultured under conditions whereby the gene is replicated during cell division, and is transcribed and translated into ubiquitin specific protease and the USP is recovered. Preferably the gene in the microorganism is recombinant DNA derived from the mRNA from cells having the active USP gene. Suitably the gene includes sequence ID No. 7, more preferably sequence ID No. 6.

According to a further aspect of the invention there is provided a new use of a protein product having Cys, QQD and His domains specified in sequence IDs numbers 2, 3 and 4, respectively, in the manufacture of a medicament for use in the diagnosis, prophylaxis or treatment of cancer. The protein has ubiquitin specific protease or ubiquitin-like specific protease activity, and has the three specified domains in the USP or UbLSP active conformation. Preferably the protein has, outside the specified domains, some level of sequence homology with sequence ID 1, for instance at least 20%, preferably at least 50%, identity with that sequence, and a level of similarity of at least 50%, preferably at least 70% or more with that sequence (in each case determined using, for instance the BLAST algorithm). Preferably the protein has sequence I.D.1 or sequence I.D.5. Homologues, such as the corresponding mouse product, described in Valero, *et al* 1999 may be used or sequences which have the above levels of identity and similarity with such a sequence.

According to a preferred embodiment the protein has sequence I.D.1.

According to another preferred embodiment the protein has sequence I.D.5.

According to a further aspect of the invention there is provided an *in vitro* method in which mammalian cells are cultured in the presence of the product of the USP25 gene located at human chromosome 21q11 or of a homologue thereof, or a functional fragment of said product.

According to a further aspect of the invention there is provided an *in vitro* method in which mammalian cells are cultured in the presence of a protein product having Cys, QQD and His domains specified in sequence IDs numbers 2, 3 and 4, respectively. Preferably the protein has, outside the specified domains, some level of sequence homology with sequence ID 1, for instance at least 20%, preferably at least 50%, identity with that sequence, and a level of similarity of at least 50%, preferably at least 70% or more with that sequence (in each case determined using, for instance the BLAST algorithm). Preferably the protein has sequence I.D.1 or sequence I.D.5. Homologues, such as the corresponding mouse product, described in Valero, *et al* 1999 may be used or sequences which have the above levels of identity and similarity with such a sequence.

In these *in vitro* methods the effect of the protein on cell growth, cell growth arrest and/or apoptosis is assessed.

The microorganism containing the specified gene has preferably been transformed with a vector comprising at least a portion of sequence I.D.6 comprising residues 199 to 3367, optionally including an additional 96 b.p. exon inserted after base 2356, and optionally including 5' and/or 3' untranslated regions (UTR's). The vector preferably comprises transcriptional and translational control sequences. The microorganism may be a yeast but is most conveniently a bacterium.

A mammalian cell transfected with a vector preferably comprises exons making up a sequence comprising residues 199 to 3367 of sequence I.D.6 and optionally also regulatory genomic sequences which are present in wild-type chromosome which lead to enhanced expression activity.

The sequence from residues 199 to 3367 of sequence ID no 6 is specified as sequence ID7 and consists of start codon (ATG), an open reading frame of 3165 nucleotides, and a stop codon (TAA).

According to a further aspect of the invention there is provided a new use of DNA including the gene located at human chromosome 21q 11-21 or a fragment thereof encoding a functional USP product in the manufacture of a medicament for use in the prophylaxis or treatment of cancer. The DNA is preferably incorporated in a gene therapy vehicle, and is preferably part of a vector, for instance a plasmid vector, or viral vector, capable of transfecting cells *in vivo*. The medicament usually includes a pharmaceutically acceptable carrier.

The invention is illustrated further in the accompanying drawings in which:

Figure 1 represents a map of human chromosome 21

Figure 2 represents a map showing the location of exon trapped products from the experiments reported below

Figure 3 represents sequence homologies of USPs

and

Figure 4 represents the results of the experiment illustrating USP properties reported below.

The following specific description describes the work which has been carried out.

Experimental

We identify a portion of human chromosome 21 homozygously deleted in non-small cell non carcinoma (NSCLC) for further study. The region contained the DNA marker with the highest NSCLC-associated loss of homozygosity (LOH), reported by Kohno *et al*. We found a shared region of overlap (SRO) for the hemizygous loss in other NSCLC. The current work is to identify genes in the SRO which have a potential role in tumour suppression.

A total of 42 fresh NSCLC cases have been analyzed from the Croatian Tumour Bank (CTB), an initiative with set rules and criteria for accumulation of fresh clinical tumor specimens for molecular studies (Spaventi *et al.*, 1994). Of these, half (including tumors #47 and #61) were samples that were recently studied for LOH of the *NM23-H1* gene (Bosnar *et al.*, 1997), and the other half were fresh tumors (data obtainable from CTB, which also lists the tumor stage in the TNM system, grade, size and survival data). In each case, the tumors and normal lung tissue specimens (as evaluated by the surgeon) were frozen in liquid nitrogen in the operating room and further stored at -70°C· Genomic DNA was isolated using standard procedures (Sambrook *et al.*, 1989). For each sample, 4μm serial frozen sections were cut, mounted on glass slides, and stained with hematoxylin-eosin (H&E). A pathologist confirmed the histologic type of the tumor and evaluated the percentage of normal cells within the tumor. Only samples with less than 20% non-tumor cells were used in this study.

Markers and LOH Analysis

Microsatellite analysis was performed using polymerase chain reaction (PCR) with appropriate primer pairs (sequences and PCR conditions as in Genome Data Base, Johns Hopkins University, Baltimore, MD), where the forward primer only from each pair was 5' fluorescently labeled with Applied Biosystems (ABI; Foster City, CA) Big Dyes™ (6-FAM or HEX). Amplification products were analysed using an ABI 310 Genetic Analyzer. Size standards (GeneScan 350) were mixed with every sample for accurate sizing; the separation of the mixture of denatured fragments was achieved by electrophoresis through a 47 cm capillary (module GS STR POP4 C) for approximately 30 min. Raw data were analyzed using GeneScan and Genotyper software. LOH ratios were calculated exactly as described in the GeneScan Applications manual provided by ABI. For each individual allele's fluorescence level, an average of 3 independent electrophoresis-analysis cycles on the ABI 310 was used for calculation.

Fluorescence In Situ Hybridization (FISH)

Unstained 4µm-thick paraffin block sections were fixed to glass slides, and a standard pretreatment protocol was followed for formalin-fixed, paraffin-embedded slides. P1-derived artificial chromosome (PAC)DNA was labelled with digoxygenin-11-dUTP (Boehringer Mannhein, Germany). Approximately 0.5 µg of each labelled PAC DNA sample was mixed with 5 µg of Cot1 DNA (Gibco BRL, Gaithersburg, MD), precipitated, denatured, allowed to preanneal, and then applied to a denatured slide and hybridized overnight. Slides were washed and signal detected using anti-digoxygenin-rhodamine, followed by DAPI counterstain. Images were captured using a Zeiss Axioskop microscope equipped with a charge-coupled device (CCD) Photometrics, Tucson, AZ) connected to an Apple Powermac 8100 computer. Images were captured on 3 levels of focus, and each level was examined for signals using SmartCapture software (Vysis, Inc., Chicago, IL). Only nuclei with signals were counted in each level, and the number of signals in each cell was determined. B: FISH using a pool of PACs 90B5, 126N20 and BAC 39112 as a probe on the paraffin embedded sections of the tumour #61. Two signal nuclei are predominant. C: FISH using a pool of PACs 73M5 and 135E14 as a probe on the paraffin embedded sections of the tumour #61. Single signal nuclei are predominant.

Northern Blot Analysis

The cDNAs were labelled by random priming and hybridized to human multiple tissue Northern blots (Clontech, Palo Alto, CA) containing 2 µg polyA + RNA per lane using the protocol recommended by the manufacturer. The exposure was for 14 hr to Molecular Dynamics (Sunnyvale, CA) Phosphorimager screens. The I.M.A.G.E. Consortium (Lennon *et al.*, 1996) cDNA clone ID 824710 and the Unigene clone A002B43 have been used as labelled probes in separate experiments.

Cleavage Analysis of

Ubiquitin-M t-β-Galactosidase Fusion Protein

This analysis was performed essentially as described (Everett *et al.*, 1997). Model fusion protein ubiquitin-Met-β-galactosidase in a pACYC184 (Cm^r replicon) was represented by the plasmid pACYC-Ub-Met-β-gal, a kind gift of R. Everett. Plasmid pRB105 containing a *Saccharomyces cerevisiae* ubiquitin-specific protease UBP2 in an IPTG-inducible pBR322 (Amp^r replicon) was a kind gift of R. Baker, and was used as a positive control. The new gene *USP25* was cloned from nucleotide position 203 to nucleotide position 3367 (numbering as in GenBank AF 134213) into *SacI/SalI* cloning sites of the IPTG-inducible *Escherichia coli* expression vector pQE30 (Qiagen, Chatsworth, CA). The *E. coli* XL-1 blue cells were transformed using a standard rubidium chloride-heat shock method with the combination of pACYC-Ub-Met-β-gal and either pQE30 vector, pQE30-*USP25*, or pRB105, and each of the 3 cotransormants was selected on medium containing chloramphenicol (42 μg/ml) and carbenicillin (75 μg/ml). Western blots were prepared by electrotransfer to a nitrocellulose membrane (Schleicher & Schull, Keene, NH). The β-galactosidase-containing bands were detected by an anti-β-galactosidase polyclonal rabit antiserum (a kind gift of R. Everett) using an enhanced chemiluminescence (ECL) assay kit (rpn 2132; Amersham, Arlington Heights, IL) under conditions recommended by the manufacturer.

Identification and cloning of USP25

Twelve sequenced exon-trapped products, when analysed using BLAST-N against public sequence databases, revealed clusters of overlapping cDNA clones. Sequences of our exon-trapped products matched exactly the sequences of the cDNAs forming contigs with a large open reading frame (ORF). In three cases (se Fig.2): from EST 824710 to AA209364, from AA307805 to AA081200 and from N92952r to Z45010, our trapped exon sequences served to bridge the gaps in the gene sequence

using PCR and suitabl restriction, ligation and chain extension techniques. The combined sequence (sequence ID no.5 and GenBank accession number AF134213) revealed a 199 bp 5'UTR, start codon, an ORF of 3165 nucleotides encoding a protein of 1055 amino acids, a stop codon, a 3'UTR of 435 nucleotides and a polyadenylation signal. The total length (without the polyA) assembled is 3803 nucleotides. On multiple human tissue Northern blots (Fig.3) a band of 4.1 kb is visible in all 16 tissues tested (including the normal human lung tissue) with a varying intensity. It is most prominent in skeletal muscle and testis, and the latter tissue also reveals a prominent shorter hybridising transcript of 1.4 kb. All tissues also show a larger weaker band of 4.9 kb, which could be due to an alternative polyadenylation site.

In the course of this analysis, the whole genomic sequence of the two PACs (73M5 and 135E14) became publicly available by the German Human Genome Sequencing Consortium (EMBL accession numbers AJ010597 and AJ010598). Comparison of the genomic sequence with the overlapping cDNA clones and exon sequences revealed that 12 out of the 24 exons had been exon-trapped (hatched rectangles in Fig.2). It also became apparent that the region immediately preceeding the first exon of the gene, comprises the known chromosome 21 CpG island at D21S382 (also known as LL56 Not I linking clone on the Not I physical map of 21q, Ichikawa et al., 1993).

When the deduced polypeptide sequence was compared to Swissprot and other public databases using BLAST-P, a clear pattern of significant homologies ($e=10^{-6}$ to 10^{-29}) to proteins across the evolutionary spectrum of eukaryotes was found (Fig.4): all of these proteins belonged to the superfamily of ubiquitin specific proteases (USP-s) or ubiquitin carboxy-terminal hydrolases (UCH-s) (Baker et al., 1992, Swanson et al., 1996, Everett et al., 1997, Wilkinson 1997, Hansen-Hagge et al., 1998, Jensen et al., 1998, Fujiwara et al., 1998, Wilkinson and Hochstrasser 1998, The C. elegans Sequencing Consortium 1998). The polypeptide sequences were most highly conserved around the three domains (the Cys box, the QQD box and the His box, Fig.4) known to be essential for the main function of these

enzymes: the cleavage of ubiquitin at its carboxy terminus from extension proteins (ubiquitin precursors) and ubiquitinated proteins and protein fragments targetted for the degradation by the 26S proteasome pathway (Wilkinson and Hochstrasser 1998). The Cysteine residue at position 178 and the Histidine residues at positions 599 and 607 (marked with an asterix in Fig.4), which were shown to be an absolute requirement for the function of USP-s and UCH-s (Amerik et al., 1997, Hansen-Hagge et al., 1998, Wilkinson and Hochstrasser 1998) were found in the correct positions in the sequence of the new gene. Since this is a first member of the USP family known to map onto human chromosome 21, we named this protein USP25, for Ubiquitin Specific Protease on Chromosome 21.

The novel protein (USP25) cleaves ubiquitin from carboxy-terminal fusion proteins

The ability of USP25 to cleave a model ubiquitin fusion protein substrate was investigated by co-expression in *E Coli.* The complete coding sequence of USP25 was cloned into a T5-driven, IPTG inducible expression vector (pQE30). The new gene USP 21 was cloned from nucleotide position 203 to nucleotide position 3367 (numbering as in sequence ID no. 2 into *Sac /Sal* cloning sites of the IPTG-inducible *E.coli* expression vector. As a positive control, the plasmid pRB105 containing a UBP2 gene encoding a *S.Cerevisiae* ubiquitin specific protease in an IPTG inducible and Amp^R vector was used. The XL-1 blue strain of *E. Coli* was co-transformed with the plasmid containing a ubiquitin-Met-β-galactosidase model fusion protein in an IPTG-inducible and chloramphenicol resistant vector, in addition to either pQE30 vector, pQE30-USP25 or the positive control (pRB105). (each of the 3 co-transformants was selected on medium containing chloramphenicol (42 μg/ml) and carbenicillin (75 μg/ml). Co-transformants were grown to exponential phase, IPTG induced, and the crude protein extracts from these cultures were analysed by Western blot using an anti β-galactosidase antibody. (The western blots were prepared by electro transfer to a nitro cellulose membrane (Schleicher and Schuel.)). The β-

galactocidase containing bands were detected by an anti-β galactosidase polyclonal rabbit anti serum using enhanced-chemiluminescence assay kit (ECL, Amersham rpn2132) under conditions recommended by the manufacturer.

As can be seen in Fig.4, the uncleaved Ub-Met-β-gal substrate (band labelled with an asterix in Fig.4, lane 4) converts to an 8 kDa shorter band (triangle in Fig.4) in the cells co-transformed with either USP25(lanes 5,6) or the yeast UBP2 expressing plasmid (lanes 9,10). Constitutive expression of USP25 (lane 5) is quite sufficient to cleave to completeness the low levels of model substrate. The more prominent and highly induced band migrating slightly further in the gel than the de-ubiquitinated cleavage product is the truncated form of b-galactosidase expressed by the XL-1 blue bacteria (compare to lanes 1,2 in Fig.4). This result demonstrates that the novel gene product named USP25 can efficiently function as a de-ubiquitinating enzyme.

From the homologies in the functional domains and from its ability to hydrolyse the bond between the C-terminal double glycine of ubiquitin and the linking methionine residue (Fig.5), it can be concluded that USP25 is a member of ubiquitin specific proteases.

Determination of Proteins with which USP25 interacts

Functional analysis of USP25 was performed with the aim of detecting the cellular proteins which interact with the USP25 protein through protein-protein interaction, using Yeast-Two-Hybrid (Y2H) approach. Saccharomyces Cerevisiae yeast has well characterised ubiquitin activating, conjugating and ligating ezymatic machinery, capable of ubiquitinating human proteins (Scheffner et al. 1998). A cDNA library from human brain cloned in "prey" vector, was co-transfected to yeast cells with USP25 cloned in "bait" vector. Since ubiquitin cleaving activity of USP25 was proven (Groet et al. 2000), this technique has a theoretical chance of detecting the natural cellular substrates for ubiquitin cleavage and de-ubiquitination by USP25. Since the action of ubiquitin cleavage is very rapid (Wilkinson and

Hochstrasser 1998), the cleavage and dissociation from its natural substrates for fully active USP25 could preclude the ability to detect the interaction through Y2H. In addition, the artificial cross-de-ubiquitination of yeast's own proteins by an overexpressed USP25 could theoretically be harmful for the yeast cell and/or for the molecular interactions required for the Y2H. For these reasons we performed Y2H using USP25-C178A, a site directed mutant we recently engineered (the mutation being of the key Cys residue in the Cys region) which abolishes the capacity for cleavage of ubiquitin by USP25, but should not interfere with the binding of USP25 to its natural ubiquitinated substrates, since this residue is conserved between all UCH-s and USP-s so far identified.

Y2H experiment using USP25-C178A cloned in the yeast two hybrid "bait" vector pAS2 (Clontech) co-transformed into yeast cells together with a human adult brain cDNA library in the "prey" vector pACT2 (Clontech). Interacting events were visualized by the activation of transcription of all three reporter genes: Ade2, Mel1 and His3. Interacting "prey" sequences were verified by PCR-sequencing on the ABI310 automated sequencer, using universal vector primers, and analysed by BLAST search on non-redundant genome and transcriptome sequence databases. The accession numbers of the sequences found to be interacting, from the GenBank database are given in the table.

Tabl 1. Summary of frequ ncy and identities of specific int racting proteins from human brain with USP25-C178A, detected using Yeast-Two-Hybrid technique

Summary of Results by decreasing frequency of detection of baits	Number of specific independent clones "fished" by Y2H	Accession number
HHR23A	8 clones	D21235
SUMO-3	8 clones	NM 006937
human UBC9	5 clones	U66867
polyUbiquitin	4 clones	AB009010
Ubiquitin	3 clones	X04803
Ran BP2 protein	1 clone	NM 006267
Various ubiquitin-like precursors (1 or 2 clones each):	4 clones	
Other proteins (1 or 2 clones each)	10 clones	

Conclusions - Interaction with HHR23A

DNA repair plays a key role in prevention of carcinogenesis and mutagenesis. This is potentially of special significance to solid tumours which have exposure to UV and chemical carcinogens as the major risk factor, such as cancers of the skin and lung. HHR23A is a homologue of yeast RAD23 protein (Masutani et al. 1994), involved in DNA excision-repair after UV damage and implicated in spindle pole body duplication and cell cycle progression in yeast (Watkins et al. 1993, Biggins et al. 1996). Human homologues HHR23A and B (Masutani et al. 1994) both belong to a group of proteins which, when mutated, lead to Xeroderma Pigmentosum, a rare autosomal recessive disorder associated with a high incidence of sunlight (UV) induced skin cancers.

More importantly, hRAD23A has also been isolated as a primary interacting protein by the same Y2H technique using E6AP as a "bait" (Kumar et al. 1999). E6AP (Human Papilloma Virus E6 associated protein)

functions as one of the two so far detect d ubiquitin ligases (attaching ubiquitin and labelling for degradation) for the master tumour suppressor p53 (Scheffner et al. 1993). The p53 and HHR23A are the only two so far proven targets for this ubiquitin ligase (Kumar et al. 1999).. Since USP25 shows high rate of target preference for HHR23A (see Table 1), and both HHR23A and p53 are ubiquitinated by E6AP, it could mean that they are both de-ubiquitinated by USP25. The fact that Y2H with USP25 did not pick up p53 is understandable, because p53 is expressed in small traces (very low level) in normal tissues, and gets only accumulated and activated following DNA damage or other stimuli for programmed cell death (apoptosis) (Haupt et al. 1997, Kubbutat et al. 1997, Lane 1998). Further experiments are therefore justified to provoke the p53 response, and monitor the effects of USP25 on p53 levels.

Moreover, lack of functional E6AP accelerates the polyglutamine-induced neuronal cell death in the mouse model for the neurodegenerative disease Spinocerebellar-ataxia 1 (SCA1) (Cummings et al. 1999). Lack of E6AP gene in a mouse expressing the polyglutamine stretch mutation of SCA1 protein dramatically reduces the presence of ubiquitinated intranuclear neuronal inclusions, but drastically accelerates the neuronal degeneration and cell death (Cummings et al. 1999). A very similar effect has been observed in Huntington's Disease (HD), where a dominant negative mutant of a ubiquitin conjugating enzyme (UBC3), when co-expressed in cultured neurons with the huntingtin protein bearing the polyglutamine extension (mutation causing HD), drastically reduces the presence of ubiquitinated intranuclear neuronal inclusions, but drastically accelerates the neuronal degeneration and cell death (Saudou et al. 1998). If USP25 de-ubiquitinates a similar set of target proteins to the ones ubiquitinated by E6AP, then the overexpression of USP25 may lead to similar effects as the inhibition of ubiquitin conjugation by E6AP. It is therefore justified to examine the effects of overexpression of USP25 (and other USP-s) on neuronal toxicity.

Interaction with human UBC9 and ubiquitin like proteins (UbL)

In yeast, there are 13 ubiquitin conjugating enzymes (E2 enzymes) (Scheffner et al. 1998). Only two (UBC3, mentioned previously in context with HD, and UBC9) are essential for cell cycle progression. Without UBC3 the cell cycle is arrested at the transition point from G1 to S phase, whereas without UBC9 the cell cycle is arrested at the transition point from G2 to M phase (Scheffner et al. 1998). Yeast UBC9, and its mammalian homologue, e.g. human UBC9 have a special function among all other UBC(E2) enzymes, in that they are specifically not conjugating to target proteins the molecule of ubiquitin, but rather of Ubiquitin-Like small proteins (UbL) (Gong et al. 1997, Schwarz et al. 1998). The yeast Smt3 and human SUMO-1 (PIC1, Sentrin, hSmt3C), SUMO-2 (hSmt3A) and SUMO-3 (hSMT3B) belong to the same family of UbL proteins with approximately 50% identity between themselves, and some 15-30% identity and 40-60% similarity in amino acid sequence to ubiquitin (Lapenta et al. 1997, Mannen et al. 1996, Kamitani et al. 1998, Saitoh and Hinchey, 2000). Yeast and human UBC9 are capable of conjugating equally yeast or human UbL-s, but not ubiquitin (Schwarz et al. 1998). The SUMO-1,-2 and -3 have the C-terminal glycine, necessary for ubiquitination of the target protein's lysine residue, but unlike ubiquitin, do not have the Lys48 residue necessary for the formation of polyubiquitin chains through isopeptide bonds, which are the signal for the proteasome degradation (Saitoh and Hinchey, 2000). Nevertheless, yeast Smt3 protein can rescue the mutant Mif2 phenotype, a deficient centromere binding protein resulting in chromosome missegregation (Meluh and Koshland 1995). SUMO-1, as well as SUMO-3 (and probably also SUMO-2) are all capable of being attached by UBC9 to RanGap1, a Ran GTP-ase activating protein (Kamitani et al. 1998). This ATP-dependent attachment is essential for the binding between modified RanGap1 and RanBP2 binding protein, in order to form functional nuclear pore complex, which controls export and import of molecules through the nuclear envelope (Mahajan et al. 1997, Matunis et al. 1998, Lee et al. 1998). In addition, UbL small proteins

have been shown to modify the death domains of Fas (Okura et al. 1996), Tumour necrosis factor receptor1 (Okura et al. 1996), PML (a tumour suppressor implicated in the pathogenesis of acute promyelocytic leukaemia) (Kamitani et al. 1998b) and Rad51/52 DNA repair proteins (Shen et al. 1996a). Their conjugating enzyme, UBC9, has been shown to interact by Y2H technique with RAD51/52 DNA repair proteins, and the master tumour suppressor p53 (Shen et al. 1996b).

The USP25 Y2H data show clear pattern of interaction in the UBC9 pathway. Interaction with UBC9 itself, is to our knowledge the first of the kind demonstration of a direct protein-protein interaction between a USP and a conjugating enzyme. Interaction with RanBP2 and SUMO-3 clearly shows that USP25 could be sharing the similar target repertoire as UBC9 (Saitoh et al. 1997). USP25 may be removing the SUMO (UbL) molecules attached to targets by UBC9. Alternatively, USP25 may be preparing the UbL-s for attachment by UBC9 to targets, by removing the oligopeptide extensions after the C-terminal Gly-Gly group from the UbL-s. The only ubiquitin protease found to be essential for yeast cell cycle progression, Ulp1, was found to be specific to Smt3 removal, and not to ubiquitin (Li and Hochstrasser 1999). This protease also had a completely different sequence from known USP-s and UCH-s. Its human homologues, SENP1 and SUSP-1, were recently cloned (Kim et al. 2000, Gong et al. 2000a), and found to specifically cleave SUMO-1,-2 and -3, but not ubiquitin and NEDD8, another UbL (Kamitani et al. 1997). The first human enzyme with classical USP structure (Cys, His domains) for which dual specificity to both ubiquitin and ubiquitin like protein was demonstrated was very recently published USP25 on chromosome 1q21 (Gong et al. 2000b). However, opposite from SENP1 and SUSP-1, this enzyme cleaves ubiquitin and Nedd8, but not SUMO-1,-2 or -3 (Gong et al. 2000b).

If USP25 cleaves the same UbL proteins to which it binds by Y2H (see Table 1), it would appear that it is the first classical USP capable of cleaving both ubiquitin and human SUMO family, which maybe very

significant to its function. Moreover, it is possible that USP25 is actually specific for SUMO-3 rather than SUMO-1 or SUMO-2. The distinction between these targets would be the first of the kind, but remains to be further confirmed. The previously observed interaction with hRAD23A (Table 1), as well as with other various precursor proteins containing the ubiquitin-like-domains (Table 1), could be linked also to its affinity to a certain type of ubiquitin like domain.

Finally, and very importantly, SUMO molecules have been shown to attach post-translationally to p53 tumour suppressor itself, by the conjugating enzyme UBC9 (Rodrigues et al. 1999, Gostissa et al. 1999). This modification for SUMO-1 in vitro requires only SUMO-1, the SUMO-1 activating enzyme and UBC9. The authors of Rodrigues et al. therefore conclude that SUMO-1 modification pathway acts as a potential regulator of the p53 response and may represent a novel target for the development of therapeutically useful modulators of the p53 response (Rodrigues et al. 1999).

We would propose that the cleavage of ubiquitin like proteins, and its association with UbC9 could be an alternative pathway explaining USP25's role in cell cycle control, and through it, its role in control of programmed cell death, with direct implications in both tumour suppression and neurodegeneration.

Figure Legends

Figure 1. Identification of the Shared Region of Overlap (S.R.O.) for hemizygous deletions in 21q11-q21 in NSCLC.A: Cytogenetic map, *Not* I long range physical map (Ishikawa, *et al.*, 1993), YAC contig (Nizetic, *et al* 1994, Shinizu *et al* 1995 and Bosch *et al* 1996), and bacterial contig, (Groet *et al* 1998) are shown in consecutive horizontal layers, respectively, above the line showing the markers used in the LOH analysis (oval symbols). Markers are named as in Genome DataBase (prefix "D21" omitted). In the column under each marker an "X" (symbolizing LOH), "+"

(an absence of LOH) and "U" (un-informative, homozygous result) for that marker in the set of eleven Croatian Tumour Bank (CTB) tumours, or in individual tumours #47 and #61 are shown. NT=not tested. For comparison with our data, markers used as probes on the genomic Southern blot of the NSCLC cell line, and/or in LOH analysis of fresh tumours in the study by Kohno and co-workers (Kohno *et al* 1998) are indicated above the empty bar symbolizing the homozygous deletion they found. In our data, hatched bars indicate hemizygous deletions, and black filled bars indicate segments showing absence of LOH or deletions. Squared symbols "X" and "+" stand for predominantly single and predominantly double signal, respectively, detected by FISH on interphase nuclei of the paraffin embedded sections of the tumour #61, when PAC clones named and indicated as bold lines in the PAC contig above the markers line, were used as probes.

Figure 2. Trapped exons (hatched rectangles) and exons deduced from overlapping sequence analysis (white rectangles) defining the exon-intron structure of the new gene USP25. Top half shows two PACs 73M5 and 135E14, also used as FISH probes in Fig. 1, which were the source of genomic DNA for exon trapping. Exon locations on the PACs are shown with vertical bars, and the 50 kbp scale bar refers to this part. Bottom half consists of overlapping cDNA fragments corresponding to exons above them, drawn in the same scale, (500 bp scale bar is shown). Names of cDNA clones are as in dB-EST and UniGene databases, 824710 is the address of the clone in the IMAGE Consortium collection. The complete cDNA sequence for the whole gene is the new GenBank entry with the accession number AF 134213.

Figure 3. Comparison of protein sequences of USP25 to other eukaryotic members of the superfamily of USP-s. The protein BAP-1 is actually from the family of Ubiquitin C-terminal Hydrolases, a distinct sub-family of this superfamily, showing homology only in the single key aminoacids in the Cys and His domains. Two reports show the localisations

of the highly homologous sequences for the HAUSP gene to 3p21 (Kashuba, *et al* 1997) and 16p13 (Robinson, *et al* 1998), respectively.

Figure 4. Demonstration of the de-ubiquitinating activity of USP25 on a model ubiquitin fusion protein. Western blot of an SDS-PAGE was detected using an anti-β-galactosidase antiserum. Lanes 1,2: the *E. coli* XL-1 blue cells alone (in all cases second line of the pair is +1PTG). Lanes 3,4: same cells co-transfected with the model fusion protein encoding plasmid pACYC-UB-Met-β-galactosidase protein, band labelled with an asterix). Lanes 5,6: as lanes 3,4 except pQE30-USP25 (full length USP25 gene cloned in the pQE30 expression vector) was added instead of pQE30. Lanes 7,8: same as lanes 3,4 except pRB105 (yeast de-ubiquitinating enzyme UBP2) was transfected instead of pQE30. Lanes 9,10: over-exposure of lanes 7,8. Note the presence of the 8kDa shorter, de-ubiquitinated Met-β-galactosidase (band labelled with a triangle).

References

Amerik, A. Yu., Swaminathan, S., Krantz, B.A., Wilkinson, K.D., and Hochstrasser, M. (1997). *In vivo* disassembly of free polyubiquitin chains by yeast Ubp14 modulates rates of protein degradation by the proteasome. *EMBO J.* 16: 4826-4838.

Baker R.T. Tobias JW, Varshavsky A. 1992. Ubiquitin-specific proteases of *Saccharomyces cerevisiae*. J. Biol Chem 267:23364-23375.

Baker, R.T., Wang, X-W., Woollatt, E., White, J.A. and Sutherlands, G.R. Identification, functional characterisation, and chromosomal localisation of USP15, a novel Human USP related to Unp Oncoprotein, and a systematic nomenclature for hUSP's. Genomics 59, 264-274 (1999).

Biggins S. et al. J. Cell Biol. 133:1331-1346 (1996)

Bosch, A., Guimera, J., Gaw, S., Gardiner, K., Chumakov, I., Patterson, D and Estivill, X. Integration of 30 CA-repeat markers into the cytogenetic, gen tic and YAC maps of human chromosome 21. Eur. J. Hum. Genet. 4, 135-142 1996.

Bosnar MH, Pav lic K, Krizanac S, Slobodnjak Z, Pavelic J. 1997. Squamous cell lung carcinomas: the role of nm23-H1 gene. J Mol Med 75:609-613.

Brodsky, G., Otterson, G.A., Parry, B.B., Hart, I., Patterson, D., and Kaye, F.J. Localization of STCH to Human Chromosome 21q11.1. Genomics, 30:627-628, 1995.

Cavailles, V., Dauvois, S., L'Horset, F., Lopez, G., Hoare, S., Kushner, P.J., and Parker, M.G. Nuclear factor RIP140 nodulates transcriptional activation by the estrogen receptor. EMBO J., 14:3741-3751, 1995.

Cummings C.J. et al., Neuron 24:879-892 (1999)

d' Andrea, A and Pellman, D. Deubiquitinating enzymes:A new class of biological regulators. Genomics 62, 395-405 1999.

Everett, R.D., Meredith, M., Orr, A., Cross, A., Kathoria, M., and Parkinson, J.A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpesvirus regularory protein. EMBO J., 16:1519-1530, 1997

Fero, M.L., Randel, E., Gurley, K.E., Roberts, J.M., and Kemp, C.J. The murine gene p27 ^{Kio1} is haplo-insufficient for tumour suppression. Nature, 396:177-180, 1998.

Fujiwara, T., Saito, A.., Suzuki, M., Shinomiya, H., Suzuki, T., Takahashi, E., Tanigami, A., Ichiyama, A., Chung, C.H., Nakamura, Y., and Tanaka, K. Identification and chromosomal assignment of USP1, a novel gene encoding a human ubiquitin-specific protease. Genomics, 54:155-158, 1998.

Gardiner K. 1996. Base composition and gene distribution: critical patterns in mammalian genome organisation. Trends Genet. 12:519-524.

Gong L. et al. J.Biol.Chem. 272(45):28198-28201 (1997)

Gong L. et al. J.Biol.Chem. 275(5):3355-3359 (2000a)

Gong L. et al. J.Biol.Chem. 275(12):14212-14216 (2000b)

Gostissa M, Hengstermann A, Fogal V, Sandy P, Schwarz SE, Scheffner M, Del Sal G (1999) Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1. EMBO J 1999 Nov 15;18(22):6462-71.

Groet, J., Ives, J.H., South A.P., Baptista, P.B., Jones, T.A., Yaspo, M-L., Lehrach, H., Potier, M-C., Van Broeckhoven, C., Nizetic, D. Bacterial contig map of the 21q11 region associated with Alzheimer's disease and abnormal myelopoiesis in Down syndrome. Genome Res., 8:385-398, 1998.

Groet, J., Ives, J.H., Jones, T.A., Danton, M., Flomen, R.H., Sheer, D., Hrasvcan, R., Pavelic, K., and Nizetic, D. Narrowing of the Region of Allelic Loss in 21q 11-21 in squamous Non-small cell Lung Carcinoma and Cloning of a Novel Ubiquitin-specific Protease Gene from the Deleted Segment. Genes Chromosomes Cancer 27:153-161 (2000).

Hansen-Hagge TE, Janssen JWG, Hameister H, Papa FR, Zechner U, Seriu T, Jauch A, Becke D, Hochstrasser M, Bartram CR. 1998. An evolutionarily conserved gene on human chromosome 5q33-q34, UBH1, encodes a novel deubiquitinating enzyme. Genomics 49:411-418.

Haupt Y, Maya R, Kazaz A, Oren M (1997) Mdm2 promotes the rapid degradation of p53, Nature 387:296-299.

Ichikawa, H., Hosoda, F., Arai, Y., Shimizu, K., Ohira, M., and Ohki, M. A *Not* I restriction map of the entire long arm of human chromosome 21. Nat. Genet., 4:361-366 1993.

Jensen DE, Proctor M, Marquis ST, Gardner HP, Ha SI, Chodosh LA, Ishov AM, Tommerup N, Vissing H, Sekido Y, Minna J, Borodovsky A, Schultz DC, Wilkinson KD, Maul GG, Barlev N, Berger SL, Predergast GC, Rauscher FJ III. 1998. BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1 - mediated cell growth suppression. Oncogene 16:1097-1112.

Kamitani T. et al. J.Biol.Chem. 272(45):28557-28562 (1997)

Kamitani T. et al. J.Biol.Chem. 273(18):11349-11353 (1998a)

Kamitani T. et al. J.Biol.Chem. 273(6):3117-3120 (1998b)

Kashuba, V.I., Grizatullin, R.Z, Protopopov AI, Allikmets R, Korolev S., Li J, Boldog R, Tory K, Zabarovska V, Marcsek Z, Sumegi J, Klein G, Zabarovsky ER, Kisselev L. 1997. NotI linking/jumping clones of human chromosome 3: mapping of the TFRC, RAB7 and HAUSP genes to regions rearranged in leukemia and deleted in solid tumours. FEBS Lett 419:181-185.

Kim I.K. et al. J.Biol.Chem. 275(19):14102-14106 (2000)

Kohno, T., Kawanishi, M., Matsuda, S., Ichikawa, H., takada, M., Ohki, M., Yammamoto, T., and Yokota. J. Homozygous deletion and frequent allelic loss of the 21q11.1-q21.1 region including the ANA gene in human lung carcinoma. Genes Chromosom. Cancer, 21:236-243, 1998.

Kubbutat M.H.G. et al. Nature 387:299-303 (1997)

Kumar S. et al. J.Biol.Chem. 274(26)18785-18792 (1999)

Lane D. 1998. Awakening angels. Nature 394:616-617.

Lapenta V. et al. Genomics 40:362-366 (1997)

Lee G.W. et al. J.Biol.Chem. 273(11):6503-6507 (1998)

Lennon GG, Auffray C. Polymeropoulos M, Soares MB. 1996. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics 33:151-152.

Li S.J. and Hochstrasser M., Nature 398:246-251 (1999)

Mahajan R. et al. Cell 88:97-107 (1997)

Mannen H. et al. Biochem.Biophys.Res.Comm. 222:178-180 (1996)

Masutani C. et al. EMBO J. 13(8):1831-1843 (1994)

Matunis M.J. et al. J.Cell Biol. 140(3):499-509 (1998)

Meluh P.B. and Koshland D., Mol. Biol. Cell 6:793-807 (1995)

Nizetic, D., Gellen L., Hamvas R., Mott R., Grigoriev A., Vatcheva R.,Zehetner G., Yaspo M.L., Dutriaux A., Lopes C., Delabar J.M., Van Broeckhoven C., Potier M.C. & Lehrach H.; An integrated YAC overlap and "cosmid-pocket" map of the human chromosome 21. Human Molecular Genetics 3(5):759-770 (1994).

Okura T. et al. J.Immunol. 157:4277-4281 (1996)

Oster, J., Mikkelsen., A. Mortality and life-tabl in Down syndrome. Acto Paediatr. Scand., 64: 322-326, 1975.

Robinson PA, Lomonte P, Leek P, Markham AF, Everett RD, 1998. Assignment of herpesvirus-associated ubiquitin-specific protease gene HAUSP to human chromosome band 16p 13.3 by in situ hybridisation, Cytogenet. Cell Genet. 83:100.

Rodriguez MS, Desterro JM, Lain S, Midgley CA, Lane DP, Hay RT (1999) SUMO-1 modification activates the transcriptional response of p53. EMBO J 1999 Nov 15;18(22):6455-61.

Saitoh H. et al. Proc.Natl.Acad.Sci.USA 94:3736-3741 (1997)

Saitoh H. and Hinchey J., J.Biol.Chem. 275(9)6252-6258 (2000)

Sambrook, J., Fritsch, E.F., & Mamiatis, T Molecular Cloning:a laboratory manual 2nd ed. Coldspring Harbor, N.Y. 1989.

Satge, D., Sasco, A.J., Geneix, A., and Malet, P. Another reason to look for tumor suppressor genes on chromosome 21. Genes Chromosomes. Cancer 21:1, 1998.

Saudou F. et al. Cell 95:55-66 (1998)

Scheffner M. et al. Cell 75:495-505 (1993)

Scheffner M. et al. The Ubiquitin-Conjugation System. In: J.M. Peters, D. Finley and R. Harris (eds.), Ubiquitin and the Biology of the Cell, Plenum Press, New York, (1998).

Schwarz S.E. et al. Proc.Natl.Acad.Sci.USA 95:560-564 (1998)

Shen Z. et al. Genomics 36:271-279 (1996a)

Shen Z. et al. Genomics 36:183-186 (1996b)

Shimizu, N., Antonarakis, S., VanBroeckhoven, C., Patterson, D., Gardiner, K., Nizetic, D., Creau, N., Delabar, J., Korenberg, J., Reeves, R., Doering, J., Ritter, O., and Cuticchia, J. Report of the fifth international workshop on human chromosome 21 mapping 1994. Cytogenet. Cell Genet., 70:147-182, 1995.

Spaventi R., Pecur, L., Pavelic, K., Pavelic, Z.P., Spaventi, S. and Sambrook, P.J. Human Tumour Bank in Croatia: as possible model for a

small bank as part of the future European Tumour Bank Network. Eur. J. Cancer 30A, 419-1994.

Swanson DA, Freund CL, Ploder L, McInnes RR, Valle D. 1996. A ubiquitin C-terminal hydrolase gene on the proximal short arm of the X chromosome: implications for X-linked retinal disorders. Hum. Mol. Genet 5:533-538.

The *C. elegans* Sequencing Consortium, 1998. Genome sequence of the nematode C. *elegans*: a platform for investigating biology. Science 282:2012-2018.

Valero, R., Martany, G., Gonzalez-Angulo, O., Gonzalez-Gonzalez, G., Puelles, L and Gonzales-Duarte, R. USP25, a Novel Gene Encoding a Deubiquitinating Enzyme, is located in the Gene-Poor Region 21q11.2. Genomics 62: 395-405 (1999).

Watkins J.F. et al. Mol. Cell Biol. 13:7757-7765 (1993)

Wilkinson, K.D. (1997). Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. *FASEB J.* 11: 1245-1256.

Wilkinson, K.D., and Hochstrasser, M. Deubiquitinating enzymes. In: J.M. Peters, D. Finley and R. Harris (eds.), Ubiquitin and the Biology of the Cell, Plenum Press, New York, 1998.

Wolozin, B. *et al* Regulation of Apoptosis by Presenilin 1. Neurobiology of Aging. Vol. 19. (15) p 523-527, 1998.

Yaspo M-L, Gellen L, Mott R, Korn B, Nizetic D, Poustka A, Lehrach H. 1995. Model for a transcript map of human chromosome 21: isolation of new coding sequences from exon and enriched cDNA libraries. Hum. Mol. Genet. 4:1291-1304.

CLAIMS

1. Use of the product of the USP25 gene located at human chromosome 21q 11-21 or a non-human homologue thereof or a functional fragment of the product manufacture of a medicament for use in the diagnosis, prophylaxis or treatment of cancer.

2. Use according to claim 1 in which the cancer is a solid tumour, for instance non-small cell lung carcinoma or skin cancer.

3. Use according to claim 1 or claim 2 in which the product is obtained from a microoganism or cell line in which the gene as recombinant DNA has been introduced in a form in which it is replicated upon cell division, and is transcribed and translated.

4. Use according to claim 3 in which the recombinant DNA is incorporated in a vector which comprises at least a portion of sequence I.D.6 comprising residues 199 to 3367, optionally including an additional 96 b.p. exon inserted after base 2356, and optionally including 5' and/or 3' untranslated regions (UTR's).

5. Use of a protein product having Cys, QQD and His domains specified in sequence IDs numbers 2, 3 and 4, respectively, in the manufacture of a medicament for use in the diagnosis, prophylaxis or treatment of cancer.

6. Use according to claim 5 in which the protein has sequence I.D. No. 1.

7. Use according to claim 5 in which the protein has sequence I.D. No. 5.

8. An *in vitro* method in which mammalian cells are cultured in the presence of a protein product having Cys, QQD and His domains specified in sequence IDs numbers 2, 3 and 4, respectively, and the effect of protein on cell growth, cell growth arrest and/or apoptosis is assessed.

9. A method according to claim 8 in which the protein has sequence I.D.1 or sequence I.D.5.

10. Use of DNA including the gene located at human chromosome 21q 11-21 (USP25) or a fragment thereof encoding a functional USP or UbLSP product in the manufacture of a medicament for use in the prophylaxis or treatment of cancer.

11. Pharmaceutical composition comprising a gene therapy vehicle and DNA including the gene located at human chromosome 21q 11-21 (USP25) or a fragment thereof encoding a functional USP or UbLSP product in transcribable form.

Figure 1

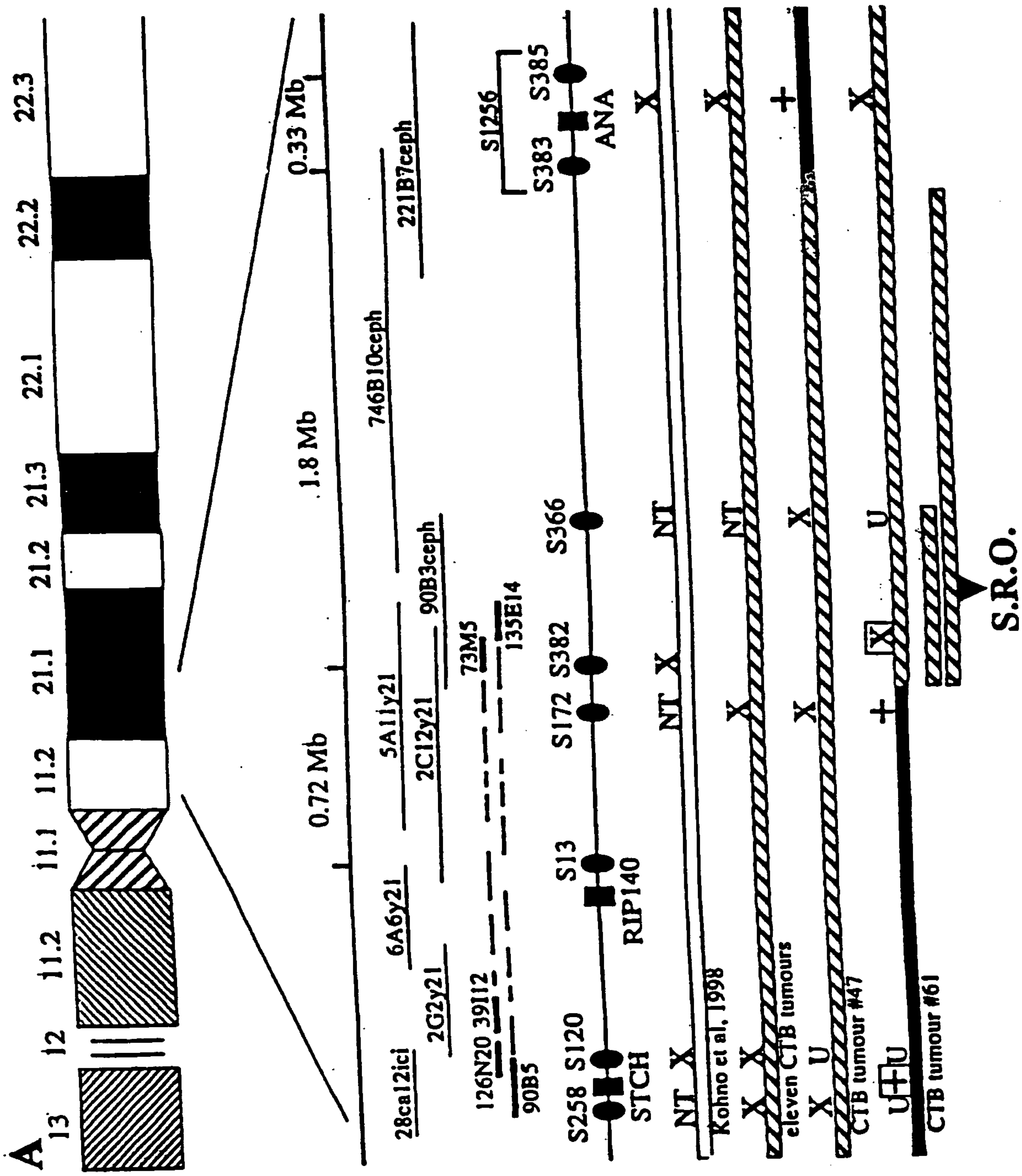

A
13
12
11.2
11.1
11.2
21.1
21.2
21.3
22.1
22.2
22.3
0.72 Mb
1.8 Mb
0.33 Mb
28ca12ici
2G2y21
6A6y21
5A11y21
2C12y21
746B10ceph
221B7ceph
90B3ceph
126N20
39I12
90B5
73M5
135E14
S258
S120
STCH
S13
RIP140
S172
S382
S366
S383
S1256
S385
ANA
Kohno et al, 1998
eleven CTB tumours
CTB tumour #47
CTB tumour #61
S.R.O.

Figure 2

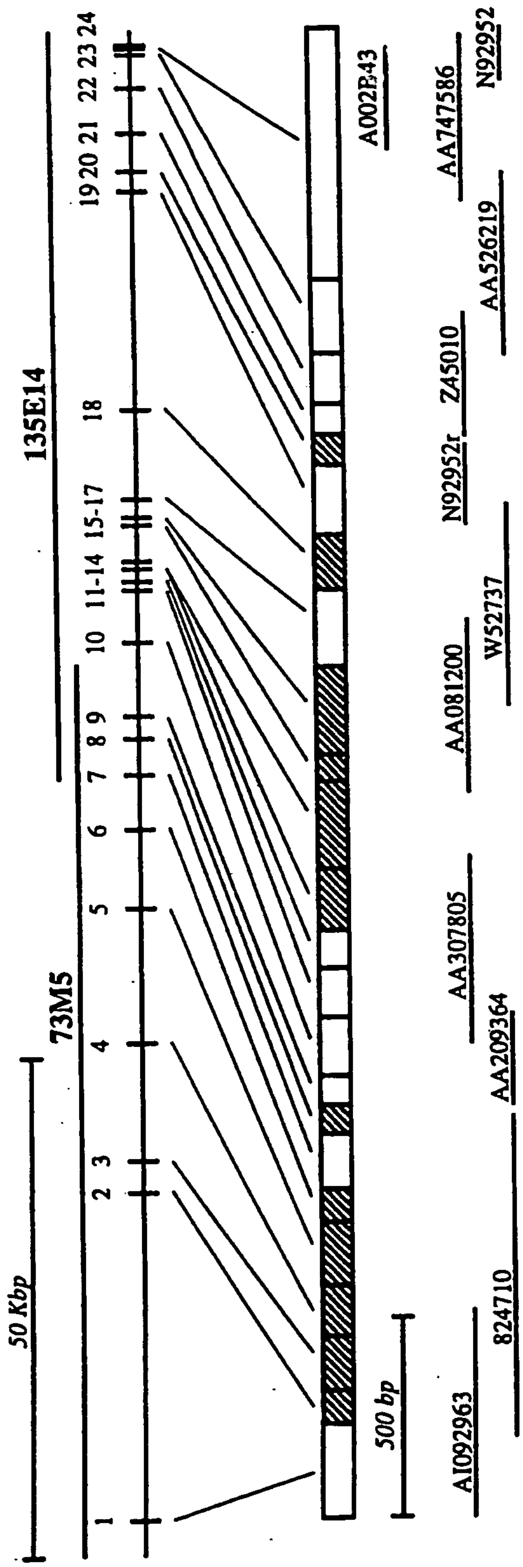

135E14
73M5
50 Kbp
500 bp
1
2 3
4
5
6
7 8 9
10
11-14
15-17
18
19 20 21 22 23 24
AA747586
N92952
AA526219
Z45010
N92952r
W52737
AA081200
AA307805
AA209364
824710
AI092963

Figure 3

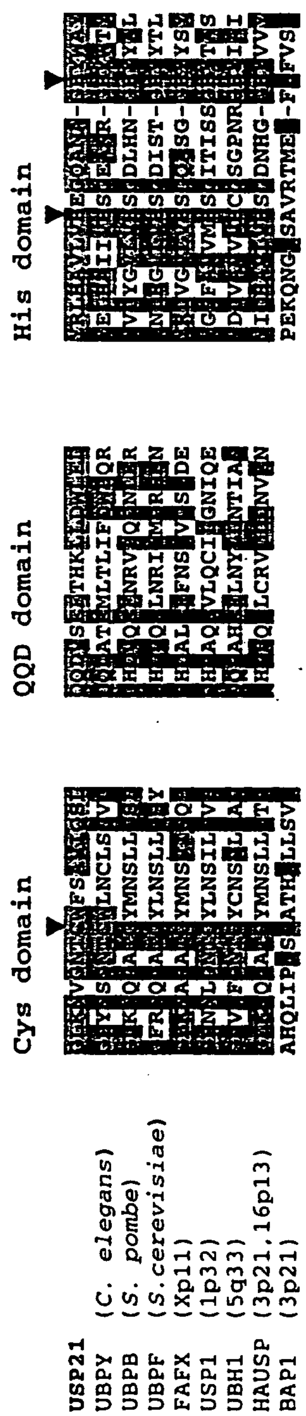
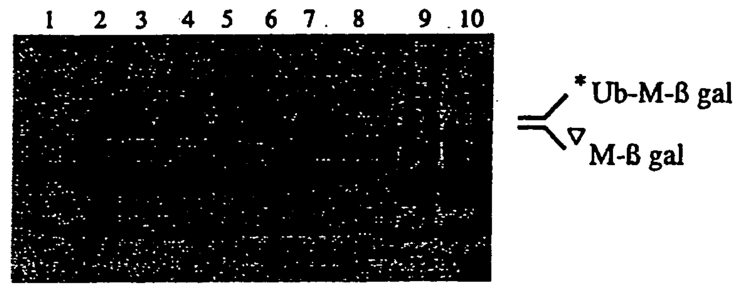

Cys domain
QQD domain
His domain
USP21
UBPY (C. elegans)
UBPB (S. pombe)
UBPF (S.cerevisiae)
FAFX (Xp11)
USP1 (1p32)
UBH1 (5q33)
HAUSP (3p21,16p13)
BAP1 (3p21)

Figure 4



Sequence Listings

SEQ ID 1 protein:

MTVEQNVLQQSAAQKHQQTFLNQLREITGINDTQILQQALKDSNGNLELAV
AFLTAKNAKTPQQEETTYYQTALPGNDRYISVGSQADTNVIDLTGDDKDDL
QRAIALSLAESNRAFRETGITDEEQAISRVLEASIAENKACLKRTPTEVWRD
SRNPYDRKRQDKAPVGLKNVGNTCWFSAVIQSLFNLLEFRRLVLNYKPPS
NAQDLPRNQKEHRNLPFMRELRYLFALLVGTKRKYVDPSRAVEILKDAFK
SNDSQQQDVSEFTHKLLDWLEDAFQMKAEEETDEEKPKNPMVELFYGRF
LAVGVLEGKKFENTEMFGQYPLQVNGFKDLHECLEAAMIEGEIESLHSEN
SGKSGQEHWFTELPPVLTFELSRFEFNQALGRPEKIHNKLEFPQVLYLDR
YMHRNREITRIKREEIKRLKDYLTVLQQRLERYLSYGSGPKRFPLVDVLQYA
LEFASSKPVCTSPVDDIDASSPPSGSIPSQTLPSTTEQQGALSSELPSTSPS
SVAAISSRSVIHKPFTQSRIPPDLPMHPAPRHITEEELSVLESCLHRWRTEIE
NDTRDLQESISRIHRTIELMYSDKSMIQVPYRLHAVLVHEGQANAGHYWAY
IFDHRESRWMKYNDIAVTKSSWEELVRDSFGGYRNASAYCLMYINDKAQFL
IQEEFNKETGQPLVGIETLPPDLRDFVEEDNQRFEKELEEWDAQLAQKALQ
EKLLASQKLRESETSVTTAQAAGDPEYLEQPSRSDFSKHLKEETIQIITKASH
EHEDKSPETVLQSAIKLEYARLVKLAQEDTPPETDYRLHHVVVYFIQNQAPK
KIIEKTLLEQFGDRNLSFDERCHNIMKVAQAKLEMIKPEEVNLEEYEEWHQD
YRKFRETTMYLIIGLENFQRESYIDSLLFLICAYQNNKELLSKGLYRGHDEELI
SHYRRECLLKLNEQAAELFESGEDREVNNGLIIMNEFIVPFLPLLLVDEMEEK
DILAVEDMRNRWCSYLGQEMEPHLQEKLTDFLPKLLDCSMEIKSFHEPPKL
PSYSTHELCERFARIMLSLSRTPADGR

SEQ ID no 2: (polypeptide)

GLKNVGNTCWFSAVIQSL

SEQ ID no 3: (polypeptide)

QQDVSEFTHKLLDWLED

SEQ ID no 4: (polypeptide)

YRLHAVLVHEGQANAGHYWAY

SEQ ID no 5: (polypeptide)

MTVEQNVLQQSAAQKHQQTFLNQLREITGINDTQILQQALKDSNGNLELA
VAFLTAKNAKTPQQEETTYYQTALPGNDRYISVGSQADTNVIDLTGDDKD
DLQRAIALSLAESNRAFRETGITDEEQAISRVLEASIAENKACLKRTPTEVW
RDSRNPYDRKRQDKAPVGLKNVGNTCWFSAVIQSLFNLLEFRRLVLNYKP
PSNAQDLPRNQKEHRNLPFMRELRYLFALLVGTKRKYVDPSRAVEILKDA
FKSNDSQQQDVSEFTHKLLDWLEDAFQMKAEEETDEEKPKNPMVELFY
GRFLAVGVLEGKKFENTEMFGQYPLQVNGFKDLHECLEAAMIEGEIESLH
SENSGKSGQEHWFTELPPVLTFELSRFEFNQALGRPEKIHNKLEFPQVLY
LDRYMHRNREITRIKREEIKRLKDYLTVLQQRLERYLSYGSGPKRFPLVDV
LQYALEFASSKPVCTSPVDDIDASSPPSGSIPSQTLPSTTEQQGALSSELP
STSPSSVAAISSRSVIHKPFTQSRIPPDLPMHPAPRHITEEKLSVLESCLHR
WRTEIENDTRDLQESISRIHRTIELMYSDKSMIQVPYRLHAVLVHEGQANA
GHYWAYIFDHRESRWMKYNDIAVTKSSWEELVRDSFGGYRNASAYCLM
YINDKAQFLIQEEFNKETGQPLVGIETLPPDLRDFVEEDNQRFEKELEEW
DAQLAQKALQEKLLASQKLRESETSVTTAQAAGDPEYLEQPSRSDFSKH
LKEETIQIITKASHEHEDKSPETVLQSIMMTPNMQGIIMAIGKSRSVYDRCG
PEAGFFKAIKLEYARLVKLAQEDTPPETDYRLHHVVVYFIQNQAPKKIIEKT
LLEQFGDRNLSFDERCHNIMKVAQAKLEMIKPEEVNLEEYEEWHQDYRK
FRETTMYLIIGLENFQRESYIDSLLFLICAYQNNKELLSKGLYRGHDEELIS
HYRRECLLKLNEQAAELFESGEDREVNNGLIIMNEFIVPFLPLLLVDEMEE
KDILAVEDMRNRWCSYLGQEMEPHLQEKLTDFLPKLLDCSMEIKSFHEP

PKLPSYSTHELCERFARIMLSLSRTPADGR

SEQ ID no 6:
ds DNA
>BASE COUNT 1223 a 752 c 841 g 987 t
>ORIGIN
> 1 acagtcggcg tttcgccgcc tgcccgcggt gcccgcgcac gccggccgcc atcgccttcg
> 61 cgcctggctg gcgggggcgc tgtcctccca ggccgtccgc gccgctccct ggagctcggc
> 121 ggagcgcggc agccagggcc ggcggaggcg cgaggagccg ggcgccaccg ccgccgccgc
> 181 cgccgccgcc gcgggggcca tgaccgtgga gcagaacgtg ctgcagcaga gcgcggcgca
> 241 gaagcaccag cagacgtttt tgaatcaact gagagaaatt acggggatta atgacaccca
> 301 gatactacag caagccttga aggatagtaa tggaaacttg gaattagcag tggctttcct
> 361 tactgcgaag aatgctaaga cccctcagca ggaggagaca acttactacc aaacagcact
> 421 tcctggcaat gatagataca tcagtgtggg aagccaagca gatacaaatg tgattgatct
> 481 cactggagat gataaagatg atcttcagag agcaattgcc ttgagtttgg ccgaatcaaa
> 541 cagggcattc agggagactg gaataactga tgaggaacaa gccattagca gagttcttga
> 601 agccagcata gcagagaata aagcatgttt gaagaggaca cctacagaag tttggaggga
> 661 ttctcgaaac ccttatgata gaaaaagaca ggacaaagct cccgttgggc taaagaatgt
> 721 tggcaatact tgttggttta gtgctgttat tcagtcatta tttaatcttt tggaatttag
> 781 aagattagtt ctgaattaca agcctccatc aaatgctcaa gatttacccc gaaaccaaaa
> 841 ggaacatcgg aatttgcctt ttatgcgtga gctgaggtat ctatttgcac ttcttgttgg
> 901 taccaaaagg aagtatgttg atccatcaag agcagttgaa attcttaagg atgctttcaa
> 961 atcaaatgac tcacagcagc aagatgtgag tgagtttaca cacaaattat tagattggtt
> 1021 agaagatgcc ttccaaatga aagctgaaga ggagacggat gaagagaagc caaagaaccc
> 1081 catggtagag ttgttctatg gcagattcct ggctgtggga gtacttgaag gtaaaaaatt
> 1141 tgaaaacact gaaatgtttg gtcagtaccc acttcaggtc aatgggttca aagatctgca
> 1201 tgagtgccta gaagctgcaa tgattgaagg agaaattgag tctttacatt cagagaattc
> 1261 aggaaaatca ggccaagagc attggtttac tgaattacca cctgtgttaa catttgaatt
> 1321 gtcaagattt gaatttaatc aggcattggg aagaccagaa aaaattcaca acaaattaga
> 1381 atttccccaa gttttatatt tggacagata catgcacaga aacagagaaa taacaagaat
> 1441 taagagggaa gagatcaaga gactgaaaga ttacctcacg gtattacaac aaaggctaga
> 1501 aagatattta agctatggtt ccggtcccaa acgattcccc ttggtagatg ttcttcagta
> 1561 tgcattggaa tttgcctcaa gtaaacctgt ttgcacttct cctgttgacg atattgacgc
> 1621 tagttcccca cctagtggtt ccataccatc acagacatta ccaagcacaa cagaacaaca
> 1681 gggagcccta tcttcagaac tgccaagcac atcaccttca tcagttgctg ccatttcatc
> 1741 gagatcagta atacacaaac catttactca gtcccggata cctccagatt tgcccatgca

> 1801 tccggcacca aggcacataa cggaggaaga actttctgtg ctggaaagtt gtttacatcg
> 1861 ctggaggaca gaaatagaaa atgacaccag agatttgcag gaaagcatat ccagaatcca
> 1921 tcgaacaatt gaattaatgt actctgacaa atctatgata caagttcctt atcgattaca
> 1981 tgccgtttta gttcacgaag gccaagctaa tgctgggcac tactgggcat atatttttga
> 2041 tcatcgtgaa agcagatgga tgaagtacaa tgatattgct gtgacaaaat catcatggga
> 2101 agagctagtg agggactctt ttggtggtta tagaaatgcc agtgcatact gtttaatgta
> 2161 cataaatgat aaggcacagt tcctaataca agaggagttt aataaagaaa ctgggcagcc
> 2221 ccttgttggt atagaaacat taccaccgga tttgagagat tttgttgagg aagacaacca
> 2281 acgatttgaa aaagaactag aagaatggga tgcacaactt gcccagaaag ctttgcagga
> 2341 aaagctttta gcgtctcaga aattgagaga gtcagagact tctgtgacaa cagcacaagc
> 2401 agcaggagac ccagaatatc tagagcagcc atcaagaagt gatttctcaa agcacttgaa
> 2461 agaagaaact attcaaataa ttaccaaggc atcacatgag catgaagata aaagtcctga
> 2521 aacagttttg cagtcggcaa ttaagttgga atatgcaagg ttggttaagt tggcccaaga
> 2581 agacaccccа ccagaaaccg attatcgttt acatcatgta gtggtctact ttatccagaa
> 2641 ccaggcacca aagaaaatta ttgagaaaac attactagaa caatttggag atagaaattt
> 2701 gagttttgat gaaaggtgtc acaacataat gaaagttgct caagccaaac tggaaatgat
> 2761 aaaacctgaa gaagtaaact tggaggaata tgaggagtgg catcaggatt ataggaaatt
> 2821 cagggaaaca actatgtatc tcataattgg gctagaaaat tttcaaagag aaagttatat
> 2881 agattccttg ctgttcctca tctgtgctta tcagaataac aaagaactct tgtctaaagg
> 2941 cttatacaga ggacatgatg aagaattgat atcacattat agaagagaat gtttgctaaa
> 3001 attaaatgag caagccgcag aactcttcga atctggagag gatcgagaag taaacaatgg
> 3061 tttgattatc atgaatgagt ttattgtccc atttttgcca ttattactgg tggatgaaat
> 3121 ggaagaaaag gatatactag ctgtagaaga tatgagaaat cgatggtgtt cctaccttgg
> 3181 tcaagaaatg gaaccacacc tccaagaaaa gctgacagat tttttgccaa aactgcttga
> 3241 ttgttctatg gagattaaaa gtttccatga gccaccgaag ttaccttcat attccacgca
> 3301 tgaactctgt gagcgatttg cccgaatcat gttgtccctc agtcgaactc ctgctgatgg
> 3361 aagataaact gcacactttc cctgaacaca ctgtataaac tctttttagt tcttaaccct
> 3421 tgccttcctg tcacagggtt tgcttgttgc tgctatagtt tttaactttt ttttatttta
> 3481 ataactgcaa aagacaaaat gactatacag actttagtca gactgcagac aataaagctg
> 3541 aaaatcgcat ggcgctcaga cattttaacc ggaactgatg tataatcaca aatctaattg
> 3601 attttattat ggcaaaacta tgcttttgcc accttcctgt tgcagtatta ctttgctttt
> 3661 atcttttctt tctcaacagc tttccattca gtctggatcc ttccatgact acagccattt
> 3721 aagtgttcag cactgtgtac gatacataat atttggtagc ttgtaaatga aataaagaat
> 3781 aaagttttat ttatggctac cta

SEQ ID no 7 dsDNA

> ca tgaccgtgga gcagaacgtg ctgcagcaga gcgcggcgca
> 241 gaagcaccag cagacgtttt tgaatcaact gagagaaatt acggggatta atgacaccca

> 301 gatactacag caagccttga aggatagtaa tggaaacttg gaattagcag tggctttcct
> 361 tactgcgaag aatgctaaga cccctcagca ggaggagaca acttactacc aaacagcact
> 421 tcctggcaat gatagataca tcagtgtggg aagccaagca gatacaaatg tgattgatct
> 481 cactggagat gataaagatg atcttcagag agcaattgcc ttgagtttgg ccgaatcaaa
> 541 cagggcattc agggagactg gaataactga tgaggaacaa gccattagca gagttcttga
> 601 agccagcata gcagagaata aagcatgttt gaagaggaca cctacagaag tttggaggga
> 661 ttctcgaaac ccttatgata gaaaaagaca ggacaaagct cccgttgggc taaagaatgt
> 721 tggcaatact tgttggttta gtgctgttat tcagtcatta tttaatcttt tggaatttag
> 781 aagattagtt ctgaattaca agcctccatc aaatgctcaa gatttacccc gaaaccaaaa
> 841 ggaacatcgg aatttgcctt ttatgcgtga gctgaggtat ctatttgcac ttcttgttgg
> 901 taccaaaagg aagtatgttg atccatcaag agcagttgaa attcttaagg atgctttcaa
> 961 atcaaatgac tcacagcagc aagatgtgag tgagtttaca cacaaattat tagattggtt
> 1021 agaagatgcc ttccaaatga aagctgaaga ggagacggat gaagagaagc caaagaaccc
> 1081 catggtagag ttgttctatg gcagattcct ggctgtggga gtacttgaag gtaaaaaatt
> 1141 tgaaaacact gaaatgtttg gtcagtaccc acttcaggtc aatgggttca aagatctgca
> 1201 tgagtgccta gaagctgcaa tgattgaagg agaaattgag tctttacatt cagagaattc
> 1261 aggaaaatca ggccaagagc attggtttac tgaattacca cctgtgttaa catttgaatt
> 1321 gtcaagattt gaatttaatc aggcattggg aagaccagaa aaaattcaca acaaattaga
> 1381 atttccccaa gttttatatt tggacagata catgcacaga aacagagaaa taacaagaat
> 1441 taagagggaa gagatcaaga gactgaaaga ttacctcacg gtattacaac aaaggctaga
> 1501 aagatattta agctatggtt ccggtcccaa acgattcccc ttggtagatg ttcttcagta
> 1561 tgcattggaa tttgcctcaa gtaaacctgt ttgcacttct cctgttgacg atattgacgc
> 1621 tagttcccca cctagtggtt ccataccatc acagacatta ccaagcacaa cagaacaaca
> 1681 gggagcccta tcttcagaac tgccaagcac atcaccttca tcagttgctg ccatttcatc
> 1741 gagatcagta atacacaaac catttactca gtcccggata cctccagatt tgcccatgca
> 1801 tccggcacca aggcacataa cggaggaaga actttctgtg ctggaaagtt gtttacatcg
> 1861 ctggaggaca gaaatagaaa atgacaccag agatttgcag gaaagcatat ccagaatcca
> 1921 tcgaacaatt gaattaatgt actctgacaa atctatgata caagttcctt atcgattaca
> 1981 tgccgtttta gttcacgaag gccaagctaa tgctgggcac tactgggcat atattttga
> 2041 tcatcgtgaa agcagatgga tgaagtacaa tgatattgct gtgacaaaat catcatggga
> 2101 agagctagtg agggactctt ttggtggtta tagaaatgcc agtgcatact gtttaatgta
> 2161 cataaatgat aaggcacagt tcctaataca agaggagttt aataaagaaa ctgggcagcc
> 2221 ccttgttggt atagaaacat taccaccgga tttgagagat tttgttgagg aagacaacca
> 2281 acgatttgaa aaagaactag aagaatggga tgcacaactt gcccagaaag ctttgcagga
> 2341 aaagctttta gcgtctcaga aattgagaga gtcagagact tctgtgacaa cagcacaagc
> 2401 agcaggagac ccagaatatc tagagcagcc atcaagaagt gatttctcaa agcacttgaa
> 2461 agaagaaact attcaaataa ttaccaaggc atcacatgag catgaagata aaagtcctga
> 2521 aacagttttg cagtcggcaa ttaagttgga atatgcaagg ttggttaagt tggcccaaga
> 2581 agacacccca ccagaaaccg attatcgttt acatcatgta gtggtctact ttatccagaa

> 2641 ccaggcacca aagaaaatta ttgagaaaac attactagaa caatttggag atagaaattt
> 2701 gagttttgat gaaaggtgtc acaacataat gaaagttgct caagccaaac tggaaatgat
> 2761 aaaacctgaa gaagtaaact tggaggaata tgaggagtgg catcaggatt ataggaaatt
> 2821 cagggaaaca actatgtatc tcataattgg gctagaaaat tttcaaagag aaagttatat
> 2881 agattccttg ctgttcctca tctgtgctta tcagaataac aaagaactct tgtctaaagg
> 2941 cttatacaga ggacatgatg aagaattgat atcacattat agaagagaat gtttgctaaa
> 3001 attaaatgag caagccgcag aactcttcga atctggagag gatcgagaag taaacaatgg
> 3061 tttgattatc atgaatgagt ttattgtccc atttttgcca ttattactgg tggatgaaat
> 3121 ggaagaaaag gatatactag ctgtagaaga tatgagaaat cgatggtgtt cctaccttgg
> 3181 tcaagaaatg gaaccacacc tccaagaaaa gctgacagat tttttgccaa aactgcttga
> 3241 ttgttctatg gagattaaaa gtttccatga gccaccgaag ttaccttcat attccacgca
> 3301 tgaactctgt gagcgatttg cccgaatcat gttgtccctc agtcgaactc ctgctgatgg
> 3361 aagataa

SEQUENCE LISTING

<110> School Of Pharmacy

<120> Tr atment of Cancer

<130> HMJ03303WO

<140> PCT/GB 0002446
<141> 2000-06-22

<150> 9914589.8
<151> 1999-06-22

<150> 0008161.2
<151> 2000-04-03

<160> 7

<170> PatentIn Ver. 2.1

<210> 1
<211> 1055
<212> PRT
<213> Homo sapiens

<400> 1

```
Met Thr Val Glu Gln Asn Val Leu Gln Gln Ser Ala Ala Gln Lys His
  1               5                  10                  15

Gln Gln Thr Phe Leu Asn Gln Leu Arg Glu Ile Thr Gly Ile Asn Asp
             20                  25                  30

Thr Gln Ile Leu Gln Gln Ala Leu Lys Asp Ser Asn Gly Asn Leu Glu
         35                  40                  45

Leu Ala Val Ala Phe Leu Thr Ala Lys Asn Ala Lys Thr Pro Gln Gln
     50                  55                  60

Glu Glu Thr Thr Tyr Tyr Gln Thr Ala Leu Pro Gly Asn Asp Arg Tyr
 65                  70                  75                  80

Ile Ser Val Gly Ser Gln Ala Asp Thr Asn Val Ile Asp Leu Thr Gly
                 85                  90                  95

Asp Asp Lys Asp Asp Leu Gln Arg Ala Ile Ala Leu Ser Leu Ala Glu
            100                 105                 110
```

Ser Asn Arg Ala Phe Arg Glu Thr Gly Ile Thr Asp Glu Glu Gln Ala
115 120 125

Ile Ser Arg Val Leu Glu Ala Ser Ile Ala Glu Asn Lys Ala Cys Leu
130 135 140

Lys Arg Thr Pro Thr Glu Val Trp Arg Asp Ser Arg Asn Pro Tyr Asp
145 150 155 160

Arg Lys Arg Gln Asp Lys Ala Pro Val Gly Leu Lys Asn Val Gly Asn
165 170 175

Thr Cys Trp Phe Ser Ala Val Ile Gln Ser Leu Phe Asn Leu Leu Glu
180 185 190

Phe Arg Arg Leu Val Leu Asn Tyr Lys Pro Pro Ser Asn Ala Gln Asp
195 200 205

Leu Pro Arg Asn Gln Lys Glu His Arg Asn Leu Pro Phe Met Arg Glu
210 215 220

Leu Arg Tyr Leu Phe Ala Leu Leu Val Gly Thr Lys Arg Lys Tyr Val
225 230 235 240

Asp Pro Ser Arg Ala Val Glu Ile Leu Lys Asp Ala Phe Lys Ser Asn
245 250 255

Asp Ser Gln Gln Gln Asp Val Ser Glu Phe Thr His Lys Leu Leu Asp
260 265 270

Trp Leu Glu Asp Ala Phe Gln Met Lys Ala Glu Glu Glu Thr Asp Glu
275 280 285

Glu Lys Pro Lys Asn Pro Met Val Glu Leu Phe Tyr Gly Arg Phe Leu
290 295 300

Ala Val Gly Val Leu Glu Gly Lys Lys Phe Glu Asn Thr Glu Met Phe
305 310 315 320

Gly Gln Tyr Pro Leu Gln Val Asn Gly Phe Lys Asp Leu His Glu Cys
325 330 335

Leu Glu Ala Ala Met Ile Glu Gly Glu Ile Glu Ser Leu His Ser Glu
340 345 350

Asn Ser Gly Lys Ser Gly Gln Glu His Trp Phe Thr Glu Leu Pro Pro
355 360 365

Val Leu Thr Phe Glu Leu Ser Arg Phe Glu Phe Asn Gln Ala Leu Gly
370 375 380

Arg Pro Glu Lys Ile His Asn Lys Leu Glu Phe Pro Gln Val Leu Tyr
385 390 395 400

Leu Asp Arg Tyr Met His Arg Asn Arg Glu Ile Thr Arg Ile Lys Arg
405 410 415

Glu Glu Ile Lys Arg Leu Lys Asp Tyr Leu Thr Val Leu Gln Gln Arg
420 425 430

Leu Glu Arg Tyr Leu Ser Tyr Gly Ser Gly Pro Lys Arg Phe Pro Leu
435 440 445

Val Asp Val Leu Gln Tyr Ala Leu Glu Phe Ala Ser Ser Lys Pro Val
450 455 460

Cys Thr Ser Pro Val Asp Asp Ile Asp Ala Ser Ser Pro Pro Ser Gly
465 470 475 480

Ser Ile Pro Ser Gln Thr Leu Pro Ser Thr Thr Glu Gln Gln Gly Ala
485 490 495

Leu Ser Ser Glu Leu Pro Ser Thr Ser Pro Ser Ser Val Ala Ala Ile
500 505 510

Ser Ser Arg Ser Val Ile His Lys Pro Phe Thr Gln Ser Arg Ile Pro
515 520 525

Pro Asp Leu Pro Met His Pro Ala Pro Arg His Ile Thr Glu Glu Glu
530 535 540

Leu Ser Val Leu Glu Ser Cys Leu His Arg Trp Arg Thr Glu Ile Glu
545 550 555 560

Asn Asp Thr Arg Asp Leu Gln Glu Ser Ile Ser Arg Ile His Arg Thr
565 570 575

Ile Glu Leu Met Tyr Ser Asp Lys Ser Met Ile Gln Val Pro Tyr Arg
580 585 590

Leu His Ala Val Leu Val His Glu Gly Gln Ala Asn Ala Gly His Tyr
595 600 605

Trp Ala Tyr Ile Phe Asp His Arg Glu Ser Arg Trp Met Lys Tyr Asn
610 615 620

```
Asp Il  Ala Val Thr Lys Ser Ser Trp Glu Glu Leu Val Arg Asp Ser
625                 630                 635                 640

Phe Gly Gly Tyr Arg Asn Ala Ser Ala Tyr Cys Leu Met Tyr Ile Asn
                645                 650                 655

Asp Lys Ala Gln Phe Leu Ile Gln Glu Glu Phe Asn Lys Glu Thr Gly
            660                 665                 670

Gln Pro Leu Val Gly Ile Glu Thr Leu Pro Pro Asp Leu Arg Asp Phe
        675                 680                 685

Val Glu Glu Asp Asn Gln Arg Phe Glu Lys Glu Leu Glu Glu Trp Asp
    690                 695                 700

Ala Gln Leu Ala Gln Lys Ala Leu Gln Glu Lys Leu Leu Ala Ser Gln
705                 710                 715                 720

Lys Leu Arg Glu Ser Glu Thr Ser Val Thr Thr Ala Gln Ala Ala Gly
                725                 730                 735

Asp Pro Glu Tyr Leu Glu Gln Pro Ser Arg Ser Asp Phe Ser Lys His
            740                 745                 750

Leu Lys Glu Glu Thr Ile Gln Ile Ile Thr Lys Ala Ser His Glu His
        755                 760                 765

Glu Asp Lys Ser Pro Glu Thr Val Leu Gln Ser Ala Ile Lys Leu Glu
    770                 775                 780

Tyr Ala Arg Leu Val Lys Leu Ala Gln Glu Asp Thr Pro Pro Glu Thr
785                 790                 795                 800

Asp Tyr Arg Leu His His Val Val Val Tyr Phe Ile Gln Asn Gln Ala
                805                 810                 815

Pro Lys Lys Ile Ile Glu Lys Thr Leu Leu Glu Gln Phe Gly Asp Arg
            820                 825                 830

Asn Leu Ser Phe Asp Glu Arg Cys His Asn Ile Met Lys Val Ala Gln
        835                 840                 845

Ala Lys Leu Glu Met Ile Lys Pro Glu Glu Val Asn Leu Glu Glu Tyr
    850                 855                 860

Glu Glu Trp His Gln Asp Tyr Arg Lys Phe Arg Glu Thr Thr Met Tyr
865                 870                 875                 880
```

Leu Ile Ile Gly Leu Glu Asn Phe Gln Arg Glu Ser Tyr Ile Asp Ser
885 890 895

Leu Leu Phe Leu Ile Cys Ala Tyr Gln Asn Asn Lys Glu Leu Leu Ser
900 905 910

Lys Gly Leu Tyr Arg Gly His Asp Glu Glu Leu Ile Ser His Tyr Arg
915 920 925

Arg Glu Cys Leu Leu Lys Leu Asn Glu Gln Ala Ala Glu Leu Phe Glu
930 935 940

Ser Gly Glu Asp Arg Glu Val Asn Asn Gly Leu Ile Ile Met Asn Glu
945 950 955 960

Phe Ile Val Pro Phe Leu Pro Leu Leu Leu Val Asp Glu Met Glu Glu
965 970 975

Lys Asp Ile Leu Ala Val Glu Asp Met Arg Asn Arg Trp Cys Ser Tyr
980 985 990

Leu Gly Gln Glu Met Glu Pro His Leu Gln Glu Lys Leu Thr Asp Phe
995 1000 1005

Leu Pro Lys Leu Leu Asp Cys Ser Met Glu Ile Lys Ser Phe His Glu
1010 1015 1020

Pro Pro Lys Leu Pro Ser Tyr Ser Thr His Glu Leu Cys Glu Arg Phe
1025 1030 1035 1040

Ala Arg Ile Met Leu Ser Leu Ser Arg Thr Pro Ala Asp Gly Arg
1045 1050 1055

<210> 2
<211> 18
<212> PRT
<213> Homo sapiens

<400> 2
Gly Leu Lys Asn Val Gly Asn Thr Cys Trp Phe Ser Ala Val Ile Gln
1 5 10 15

Ser Leu

<210> 3
<211> 17
<212> PRT
<213> Homo sapiens

<400> 3
Gln Gln Asp Val Ser Glu Phe Thr His Lys Leu Leu Asp Trp Leu Glu
1   5   10   15

Asp

<210> 4
<211> 21
<212> PRT
<213> Homo sapiens

<400> 4
Tyr Arg Leu His Ala Val Leu Val His Glu Gly Gln Ala Asn Ala Gly
1   5   10   15

His Tyr Trp Ala Tyr
            20

<210> 5
<211> 1087
<212> PRT
<213> Homo sapiens

<400> 5
Met Thr Val Glu Gln Asn Val Leu Gln Gln Ser Ala Ala Gln Lys His
1   5   10   15

Gln Gln Thr Phe Leu Asn Gln Leu Arg Glu Ile Thr Gly Ile Asn Asp
            20   25   30

Thr Gln Ile Leu Gln Gln Ala Leu Lys Asp Ser Asn Gly Asn Leu Glu
        35   40   45

Leu Ala Val Ala Phe Leu Thr Ala Lys Asn Ala Lys Thr Pro Gln Gln
    50   55   60

Glu Glu Thr Thr Tyr Tyr Gln Thr Ala Leu Pro Gly Asn Asp Arg Tyr

65 70 75 80

Ile Ser Val Gly Ser Gln Ala Asp Thr Asn Val Ile Asp Leu Thr Gly
85 90 95

Asp Asp Lys Asp Asp Leu Gln Arg Ala Ile Ala Leu Ser Leu Ala Glu
100 105 110

Ser Asn Arg Ala Phe Arg Glu Thr Gly Ile Thr Asp Glu Glu Gln Ala
115 120 125

Ile Ser Arg Val Leu Glu Ala Ser Ile Ala Glu Asn Lys Ala Cys Leu
130 135 140

Lys Arg Thr Pro Thr Glu Val Trp Arg Asp Ser Arg Asn Pro Tyr Asp
145 150 155 160

Arg Lys Arg Gln Asp Lys Ala Pro Val Gly Leu Lys Asn Val Gly Asn
165 170 175

Thr Cys Trp Phe Ser Ala Val Ile Gln Ser Leu Phe Asn Leu Leu Glu
180 185 190

Phe Arg Arg Leu Val Leu Asn Tyr Lys Pro Pro Ser Asn Ala Gln Asp
195 200 205

Leu Pro Arg Asn Gln Lys Glu His Arg Asn Leu Pro Phe Met Arg Glu
210 215 220

Leu Arg Tyr Leu Phe Ala Leu Leu Val Gly Thr Lys Arg Lys Tyr Val
225 230 235 240

Asp Pro Ser Arg Ala Val Glu Ile Leu Lys Asp Ala Phe Lys Ser Asn
245 250 255

Asp Ser Gln Gln Gln Asp Val Ser Glu Phe Thr His Lys Leu Leu Asp
260 265 270

Trp Leu Glu Asp Ala Phe Gln Met Lys Ala Glu Glu Glu Thr Asp Glu
275 280 285

Glu Lys Pro Lys Asn Pro Met Val Glu Leu Phe Tyr Gly Arg Phe Leu
290 295 300

Ala Val Gly Val Leu Glu Gly Lys Lys Phe Glu Asn Thr Glu Met Phe
305 310 315 320

Gly Gln Tyr Pro Leu Gln Val Asn Gly Phe Lys Asp Leu His Glu Cys

325 330 335

Leu Glu Ala Ala Met Ile Glu Gly Glu Ile Glu Ser Leu His Ser Glu
340 345 350

Asn Ser Gly Lys Ser Gly Gln Glu His Trp Phe Thr Glu Leu Pro Pro
355 360 365

Val Leu Thr Phe Glu Leu Ser Arg Phe Glu Phe Asn Gln Ala Leu Gly
370 375 380

Arg Pro Glu Lys Ile His Asn Lys Leu Glu Phe Pro Gln Val Leu Tyr
385 390 395 400

Leu Asp Arg Tyr Met His Arg Asn Arg Glu Ile Thr Arg Ile Lys Arg
405 410 415

Glu Glu Ile Lys Arg Leu Lys Asp Tyr Leu Thr Val Leu Gln Gln Arg
420 425 430

Leu Glu Arg Tyr Leu Ser Tyr Gly Ser Gly Pro Lys Arg Phe Pro Leu
435 440 445

Val Asp Val Leu Gln Tyr Ala Leu Glu Phe Ala Ser Ser Lys Pro Val
450 455 460

Cys Thr Ser Pro Val Asp Asp Ile Asp Ala Ser Ser Pro Pro Ser Gly
465 470 475 480

Ser Ile Pro Ser Gln Thr Leu Pro Ser Thr Thr Glu Gln Gln Gly Ala
485 490 495

Leu Ser Ser Glu Leu Pro Ser Thr Ser Pro Ser Ser Val Ala Ala Ile
500 505 510

Ser Ser Arg Ser Val Ile His Lys Pro Phe Thr Gln Ser Arg Ile Pro
515 520 525

Pro Asp Leu Pro Met His Pro Ala Pro Arg His Ile Thr Glu Glu Lys
530 535 540

Leu Ser Val Leu Glu Ser Cys Leu His Arg Trp Arg Thr Glu Ile Glu
545 550 555 560

Asn Asp Thr Arg Asp Leu Gln Glu Ser Ile Ser Arg Ile His Arg Thr
565 570 575

Ile Glu Leu Met Tyr Ser Asp Lys Ser Met Ile Gln Val Pro Tyr Arg

580 585 590

Leu His Ala Val Leu Val His Glu Gly Gln Ala Asn Ala Gly His Tyr
595 600 605

Trp Ala Tyr Ile Phe Asp His Arg Glu Ser Arg Trp Met Lys Tyr Asn
610 615 620

Asp Ile Ala Val Thr Lys Ser Ser Trp Glu Glu Leu Val Arg Asp Ser
625 630 635 640

Phe Gly Gly Tyr Arg Asn Ala Ser Ala Tyr Cys Leu Met Tyr Ile Asn
645 650 655

Asp Lys Ala Gln Phe Leu Ile Gln Glu Glu Phe Asn Lys Glu Thr Gly
660 665 670

Gln Pro Leu Val Gly Ile Glu Thr Leu Pro Pro Asp Leu Arg Asp Phe
675 680 685

Val Glu Glu Asp Asn Gln Arg Phe Glu Lys Glu Leu Glu Glu Trp Asp
690 695 700

Ala Gln Leu Ala Gln Lys Ala Leu Gln Glu Lys Leu Leu Ala Ser Gln
705 710 715 720

Lys Leu Arg Glu Ser Glu Thr Ser Val Thr Thr Ala Gln Ala Ala Gly
725 730 735

Asp Pro Glu Tyr Leu Glu Gln Pro Ser Arg Ser Asp Phe Ser Lys His
740 745 750

Leu Lys Glu Glu Thr Ile Gln Ile Ile Thr Lys Ala Ser His Glu His
755 760 765

Glu Asp Lys Ser Pro Glu Thr Val Leu Gln Ser Ile Met Met Thr Pro
770 775 780

Asn Met Gln Gly Ile Ile Met Ala Ile Gly Lys Ser Arg Ser Val Tyr
785 790 795 800

Asp Arg Cys Gly Pro Glu Ala Gly Phe Phe Lys Ala Ile Lys Leu Glu
805 810 815

Tyr Ala Arg Leu Val Lys Leu Ala Gln Glu Asp Thr Pro Pro Glu Thr
820 825 830

Asp Tyr Arg Leu His His Val Val Val Tyr Phe Ile Gln Asn Gln Ala

835 840 845

Pro Lys Lys Ile Ile Glu Lys Thr Leu Leu Glu Gln Phe Gly Asp Arg
850 855 860

Asn Leu Ser Phe Asp Glu Arg Cys His Asn Ile Met Lys Val Ala Gln
865 870 875 880

Ala Lys Leu Glu Met Ile Lys Pro Glu Glu Val Asn Leu Glu Glu Tyr
885 890 895

Glu Glu Trp His Gln Asp Tyr Arg Lys Phe Arg Glu Thr Thr Met Tyr
900 905 910

Leu Ile Ile Gly Leu Glu Asn Phe Gln Arg Glu Ser Tyr Ile Asp Ser
915 920 925

Leu Leu Phe Leu Ile Cys Ala Tyr Gln Asn Asn Lys Glu Leu Leu Ser
930 935 940

Lys Gly Leu Tyr Arg Gly His Asp Glu Glu Leu Ile Ser His Tyr Arg
945 950 955 960

Arg Glu Cys Leu Leu Lys Leu Asn Glu Gln Ala Ala Glu Leu Phe Glu
965 970 975

Ser Gly Glu Asp Arg Glu Val Asn Asn Gly Leu Ile Ile Met Asn Glu
980 985 990

Phe Ile Val Pro Phe Leu Pro Leu Leu Leu Val Asp Glu Met Glu Glu
995 1000 1005

Lys Asp Ile Leu Ala Val Glu Asp Met Arg Asn Arg Trp Cys Ser Tyr
1010 1015 1020

Leu Gly Gln Glu Met Glu Pro His Leu Gln Glu Lys Leu Thr Asp Phe
1025 1030 1035 1040

Leu Pro Lys Leu Leu Asp Cys Ser Met Glu Ile Lys Ser Phe His Glu
1045 1050 1055

Pro Pro Lys Leu Pro Ser Tyr Ser Thr His Glu Leu Cys Glu Arg Phe
1060 1065 1070

Ala Arg Ile Met Leu Ser Leu Ser Arg Thr Pro Ala Asp Gly Arg
1075 1080 1085

<210> 6
<211> 3803
<212> DNA
<213> Homo sapiens

<400> 6
acagtcggcg tttcgc[illegible]gcc tgcccgcggt gcccgcgcac gccggccgcc atcgccttcg 60
cgcctggctg gcgggggcgc tgtcctccca ggccgtccgc gccgctccct ggagctcggc 120
ggagcgcggc agccagggcc ggcggaggcg cgaggagccg ggcgccaccg ccgccgccgc 180
cgccgccgcc gcgggggcca tgaccgtgga gcagaacgtg ctgcagcaga gcgcggcgca 240
gaagcaccag cagacgtttt tgaatcaact gagagaaatt acggggatta atgacaccca 300
gatactacag caagccttga aggatagtaa tggaaacttg gaattagcag tggctttcct 360
tactgcgaag aatgctaaga cccctcagca ggaggagaca acttactacc aaacagcact 420
tcctggcaat gatagataca tcagtgtggg aagccaagca gatacaaatg tgattgatct 480
cactggagat gataaagatg atcttcagag agcaattgcc ttgagtttgg ccgaatcaaa 540
cagggcattc agggagactg gaataactga tgaggaacaa gccattagca gagttcttga 600
agccagcata gcagagaata aagcatgttt gaagaggaca cctacagaag tttggaggga 660
ttctcgaaac ccttatgata gaaaaagaca ggacaaagct cccgttgggc taaagaatgt 720
tggcaatact tgttggttta gtgctgttat tcagtcatta tttaatcttt tggaatttag 780
aagattagtt ctgaattaca agcctccatc aaatgctcaa gatttacccc gaaaccaaaa 840
ggaacatcgg aatttgcctt ttatgcgtga gctgaggtat ctatttgcac ttcttgttgg 900
taccaaaagg aagtatgttg atccatcaag agcagttgaa attcttaagg atgctttcaa 960
atcaaatgac tcacagcagc aagatgtgag tgagtttaca cacaaattat tagattggtt 1020
agaagatgcc ttccaaatga aagctgaaga ggagacggat gaagagaagc caaagaaccc 1080
catggtagag ttgttctatg gcagattcct ggctgtggga gtacttgaag gtaaaaaatt 1140
tgaaaacact gaaatgtttg gtcagtaccc acttcaggtc aatgggttca aagatctgca 1200
tgagtgccta gaagctgcaa tgattgaagg agaaattgag tctttacatt cagagaattc 1260
aggaaaatca ggccaagagc attggtttac tgaattacca cctgtgttaa catttgaatt 1320
gtcaagattt gaatttaatc aggcattggg aagaccagaa aaaattcaca acaaattaga 1380
atttccccaa gttttatatt tggacagata catgcacaga aacagagaaa taacaagaat 1440
taagagggaa gagatcaaga gactgaaaga ttacctcacg gtattacaac aaaggctaga 1500
aagatattta agctatggtt ccggtcccaa acgattcccc ttggtagatg ttcttcagta 1560
tgcattggaa tttgcctcaa gtaaacctgt ttgcacttct cctgttgacg atattgacgc 1620
tagttcccca cctagtggtt ccataccatc acagacatta ccaagcacaa cagaacaaca 1680
gggagcccta tcttcagaac tgccaagcac atcaccttca tcagttgctg ccatttcatc 1740
gagatcagta atacacaaac catttactca gtcccggata cctccagatt tgcccatgca 1800
tccggcacca aggcacataa cggaggaaga actttctgtg ctggaaagtt gtttacatcg 1860
ctggaggaca gaaatagaaa atgacaccag agatttgcag gaaagcatat ccagaatcca 1920
tcgaacaatt gaattaatgt actctgacaa atctatgata caagttcctt atcgattaca 1980
tgccgtttta gttcacgaag gccaagctaa tgctgggcac tactgggcat atatttttga 2040
tcatcgtgaa agcagatgga tgaagtacaa tgatattgct gtgacaaaat catcatggga 2100
agagctagtg agggactctt ttggtggtta tagaaatgcc agtgcatact gtttaatgta 2160
cataaatgat aaggcacagt tcctaataca agaggagttt aataaagaaa ctgggcagcc 2220
ccttgttggt atagaaacat taccaccgga tttgagagat tttgttgagg aagacaacca 2280
acgatttgaa aaagaactag aagaatggga tgcacaactt gcccagaaag ctttgcagga 2340
aaagctttta gcgtctcaga aattgagaga gtcagagact tctgtgacaa cagcacaagc 2400
agcaggagac ccagaatatc tagagcagcc atcaagaagt gatttctcaa agcacttgaa 2460

agaagaaact attcaaataa ttaccaaggc atcacatgag catgaagata aaagtcctga 2520
aacagttttg cagtcggcaa ttaagttgga atatgcaagg ttggttaagt tggcccaaga 2580
agacacccca ccagaaaccg attatcgttt acatcatgta gtggtctact ttatccagaa 2640
ccaggcacca aagaaaatta ttgagaaaac attactagaa caatttggag atagaaattt 2700
gagttttgat gaaaggtgtc acaacataat gaaagttgct caagccaaac tggaaatgat 2760
aaaacctgaa gaagtaaact tggaggaata tgaggagtgg catcaggatt ataggaaatt 2820
cagggaaaca actatgtatc tcataattgg gctagaaaat tttcaaagag aaagttatat 2880
agattccttg ctgttcctca tctgtgctta tcagaataac aaagaactct tgtctaaagg 2940
cttatacaga ggacatgatg aagaattgat atcacattat agaagagaat gtttgctaaa 3000
attaaatgag caagccgcag aactcttcga atctggagag gatcgagaag taaacaatgg 3060
tttgattatc atgaatgagt ttattgtccc atttttgcca ttattactgg tggatgaaat 3120
ggaagaaaag gatatactag ctgtagaaga tatgagaaat cgatggtgtt cctaccttgg 3180
tcaagaaatg gaaccacacc tccaagaaaa gctgacagat tttttgccaa aactgcttga 3240
ttgttctatg gagattaaaa gtttccatga gccaccgaag ttaccttcat attccacgca 3300
tgaactctgt gagcgatttg cccgaatcat gttgtccctc agtcgaactc ctgctgatgg 3360
aagataaact gcacactttc cctgaacaca ctgtataaac tctttttagt tcttaaccct 3420
tgccttcctg tcacagggtt tgcttgttgc tgctatagtt tttaactttt ttttatttta 3480
ataactgcaa aagacaaaat gactatacag actttagtca gactgcagac aataaagctg 3540
aaaatcgcat ggcgctcaga cattttaacc ggaactgatg tataatcaca aatctaattg 3600
attttattat ggcaaaacta tgcttttgcc accttcctgt tgcagtatta ctttgctttt 3660
atcttttctt tctcaacagc tttccattca gtctggatcc ttccatgact acagccattt 3720
aagtgttcag cactgtgtac gatacataat atttggtagc ttgtaaatga aataaagaat 3780
aaagttttat ttatggctac cta 3803

<210> 7
<211> 3169
<212> DNA
<213> Homo sapiens

<400> 7
catgaccgtg gagcagaacg tgctgcagca gagcgcggcg cagaagcacc agcagacgtt 60
tttgaatcaa ctgagagaaa ttacggggat taatgacacc cagatactac agcaagcctt 120
gaaggatagt aatggaaact tggaattagc agtggctttc cttactgcga agaatgctaa 180
gacccctcag caggaggaga caacttacta ccaaacagca cttcctggca atgatagata 240
catcagtgtg ggaagccaag cagatacaaa tgtgattgat ctcactggag atgataaaga 300
tgatcttcag agagcaattg ccttgagttt ggccgaatca aacagggcat tcagggagac 360
tggaataact gatgaggaac aagccattag cagagttctt gaagccagca tagcagagaa 420
taaagcatgt ttgaagagga cacctacaga agtttggagg gattctcgaa accccttatga 480
tagaaaaaga caggacaaag ctcccgttgg gctaaagaat gttggcaata cttgttggtt 540
tagtgctgtt attcagtcat tatttaatct tttggaattt agaagattag ttctgaatta 600
caagcctcca tcaaatgctc aagatttacc ccgaaaccaa aaggaacatc ggaatttgcc 660
ttttatgcgt gagctgaggt atctatttgc acttcttgtt ggtaccaaaa ggaagtatgt 720
tgatccatca agagcagttg aaattcttaa ggatgctttc aaatcaaatg actcacagca 780
gcaagatgtg agtgagttta cacacaaatt attagattgg ttagaagatg ccttccaaat 840
gaaagctgaa gaggagacgg atgaagagaa gccaaagaac cccatggtag agttgttcta 900
tggcagattc ctggctgtgg gagtacttga aggtaaaaaa tttgaaaaca ctgaaatgtt 960
tggtcagtac ccacttcagg tcaatgggtt caaagatctg catgagtgcc tagaagctgc 1020

```
aatgattgaa ggagaaattg agtctttaca ttcagagaat tcaggaaaat caggccaaga 1080
gcattggttt actgaattac cacctgtgtt aacatttgaa ttgtcaagat ttgaatttaa 1140
tcaggcattg ggaagaccag aaaaaattca caacaaatta gaatttcccc aagttttata 1200
tttggacaga tacatgcaca gaaacagaga aataacaaga attaagaggg aagagatcaa 1260
gagactgaaa gattacctca cggtattaca acaaaggcta gaaagatatt taagctatgg 1320
ttccggtccc aaacgattcc ccttggtaga tgttcttcag tatgcattgg aatttgcctc 1380
aagtaaacct gtttgcactt ctcctgttga cgatattgac gctagttccc cacctagtgg 1440
ttccatacca tcacagacat taccaagcac aacagaacaa cagggagccc tatcttcaga 1500
actgccaagc acatcacctt catcagttgc tgccatttca tcgagatcag taatacacaa 1560
accatttact cagtcccgga tacctccaga tttgcccatg catccggcac caaggcacat 1620
aacggaggaa gaactttctg tgctggaaag ttgtttacat cgctggagga cagaaataga 1680
aaatgacacc agagatttgc aggaaagcat atccagaatc catcgaacaa ttgaattaat 1740
gtactctgac aaatctatga tacaagttcc ttatcgatta catgccgttt tagttcacga 1800
aggccaagct aatgctgggc actactgggc atatattttt gatcatcgtg aaagcagatg 1860
gatgaagtac aatgatattg ctgtgacaaa atcatcatgg gaagagctag tgagggactc 1920
ttttggtggt tatagaaatg ccagtgcata ctgtttaatg tacataaatg ataaggcaca 1980
gttcctaata caagaggagt ttaataaaga aactgggcag ccccttgttg gtatagaaac 2040
attaccaccg gatttgagag attttgttga ggaagacaac caacgatttg aaaaagaact 2100
agaagaatgg gatgcacaac ttgcccagaa agctttgcag gaaaagcttt tagcgtctca 2160
gaaattgaga gagtcagaga cttctgtgac aacagcacaa gcagcaggag acccagaata 2220
tctagagcag ccatcaagaa gtgatttctc aaagcacttg aaagaagaaa ctattcaaat 2280
aattaccaag gcatcacatg agcatgaaga taaaagtcct gaaacagttt tgcagtcggc 2340
aattaagttg gaatatgcaa ggttggttaa gttggcccaa gaagacaccc caccagaaac 2400
cgattatcgt ttacatcatg tagtggtcta ctttatccag aaccaggcac caaagaaaat 2460
tattgagaaa acattactag aacaatttgg agatagaaat ttgagttttg atgaaaggtg 2520
tcacaacata atgaaagttg ctcaagccaa actggaaatg ataaaacctg aagaagtaaa 2580
cttggaggaa tatgaggagt ggcatcagga ttataggaaa ttcagggaaa caactatgta 2640
tctcataatt gggctagaaa attttcaaag agaaagttat atagattcct tgctgttcct 2700
catctgtgct tatcagaata acaaagaact cttgtctaaa ggcttataca gaggacatga 2760
tgaagaattg atatcacatt atagaagaga atgtttgcta aaattaaatg agcaagccgc 2820
agaactcttc gaatctggag aggatcgaga agtaaacaat ggtttgatta tcatgaatga 2880
gtttattgtc ccatttttgc cattattact ggtggatgaa atggaagaaa aggatatact 2940
agctgtagaa gatatgagaa atcgatggtg ttcctacctt ggtcaagaaa tggaaccaca 3000
cctccaagaa aagctgacag attttttgcc aaaactgctt gattgttcta tggagattaa 3060
aagtttccat gagccaccga agttaccttc atattccacg catgaactct gtgagcgatt 3120
tgcccgaatc atgttgtccc tcagtcgaac tcctgctgat ggaagataa            3169
```